Proyecto Final de Carrera

Sistema estereoscópico para teleoperación asistida y supervisión de tareas robotizadas

Ingrid Capell Muñoz

Octubre 2006

Escola Superior d'Enginyeria Industrial de Barcelona

1 Resumen

El objetivo de este proyecto es diseñar un sistema de visión estereoscópica con cliente remoto para la supervisión de tareas robotizadas dentro del proyecto "Teleoperación asistida y supervisión de tareas robotizadas a través de Internet 2"(TASTRI2).

El sistema consta de dos módulos. Un primer módulo sensor está compuesto de dos cámaras Canon VC-C5 en paralelo, que forman parte de una arquitectura cliente-servidor remoto. Esta arquitectura permite el control de las dos cámaras gracias, por un lado, a la creación de una Librería de control de las cámaras, y, por otro, al desarrollo de un cliente Qt de las cámaras y a un servidor remoto que se comunica vía RS-232 con las mismas. El segundo módulo desarrollado en este proyecto se encarga del tratamiento y posterior procesado de las imágenes que se extraen de estas cámaras . La parte de tratamiento consiste, a su vez, por un lado, en la compresión por hard de estas imágenes (en formato MPEG4), y por otro, en su correspondiente captura con un servidor de RTSP, en este caso Spook, y su envío a través de la red de Internet. Finalmente, para el procesado de las imágenes se ha desarrollado un programa que, en primer lugar, obtiene paquetes del stream de video, los trata y, una vez decodificados, obtiene las imágenes que componen dicho vídeo; en segundo lugar, se muestran las imágenes obtenidas en una ventana gráfica. Dicha ventana, gracias a la tarjeta de vídeo, capaz de soportar estereoscopía, y a la sincronización del quad buffer del PC con unas gafas con obturación (shutter glasses), permite la correcta visualización en 3D de las imágenes procedentes de las cámaras.

La memoria del proyecto consta de una primera parte introductoria al mundo de la estereoscopía donde se definen aquellos conceptos más relevantes para la comprensión de las diferentes fases del proyecto. En segundo lugar, se realiza un estudio de las características y de cuáles son los condicionantes de un sistema estereoscópico para, en tercer lugar, poder plantear el diseño y realización de los dos módulos anteriormente mencionados. En cuarto lugar se procede a evaluar los resultados de dicho sistema, a analizar económicamente los costes de viabilidad del proyecto y a describir el impacto ambiental del mismo. Cabe destacar, que estos dos últimos conceptos se desarrollan en los anexos de esta memoria. Finalmente se exponen las conclusiones del proyecto y se plantea el posible trabajo futuro de este sistema estereoscópico.

Índice

Índice de figuras

2 Glosario

Glosario de la terminología inglesa empleada en los papers i bibliografía consultada con posibles traducciónes y definiciones en castellano.

API: Del inglés *Application Programming Interface* - Interfaz de Programación de Aplicaciones, es un conjunto de especificaciones de comunicación entre componentes software. Representa un método para conseguir abstracción en la programación y proporcionar un conjunto de funciones de uso general de forma que los programadores hagan uso de éstas evitándose el trabajo de reprogramarlo todo desde el principio.

Bit-rate: Del inglés *Flujo de bits* Datos por segundo que contiene un vídeo. Mide la velocidad. Se entiende que a mayor bit-rate, mayor calidad de imagen, aunque mayor ancho de banda necesario para su correcta reproducción. Si se observan saltos en la reproducción de un vídeo es muy posible que se deba a que el PC no puede mantener una velocidad constante de reproducción. El buen uso del bit-rate así como la correcta elección del codec son determinantes en la calidad de un vídeo.

Byte: Unidad básica de almacenamiento de información, equivale a ocho bits. En español el equivalente para este anglicismo es octeto, si bien la Real Academia Española ha aceptado el término.

CCD: Del inglés *Charge Couple Device* es un dispositivo de acoplamiento de carga constituido por una matriz lineal o por una bidimensional de elementos sensibles a la luz. La luz se convierte en una carga eléctrica proporcional a la luz que incide en cada célula. Las células están acopladas a un sistema de barrido que, después de una conversión de analógico a digital, presenta la imagen como una serie de dígitos binarios.

Codec: *Compresor-decompresor*, y es una extensión instalada en el sistema que puede servir para codificar vídeo a un formato determinado, así como decodificarlo o reproducirlo. La diversidad de codecs existentes está ligada a los diferentes bit-rates, así como al medio de salida al que estén destinados. Todos buscan la mejor calidad de imagen al menor bit-rate posible. Hay codecs denominados lossless o «sin pérdidas», que conservan toda la calidad de imagen del vídeo original. Esto lo hacen a costa de un bit-rate alto, por lo cual suelen

utilizarse para «transportar» un clip de vídeo entre diferentes aplicaciones y/o plataformas en el proceso de postproducción sin merma de calidad.

Codec y formato contenedor: Un codec es un algoritmo de compresión, utilizado para reducir el tamaño de un flujo. Existen codecs de audio y codecs de vídeo. MPEG-1, MPEG-2, MPEG-4, Vorbis, DivX, ... son codecs. Sin embargo, un formato contenedor contiene uno o varios flujos ya codificados por codecs. A menudo, hay un flujo de audio y uno de vídeo. AVI, Ogg, MOV, ASF, ... son formatos contenedor. Los flujos que contengan pueden ser codificados utilizando diferentes codecs. En un caso ideal, se podría utilizar cualquier codec en cualquier formato contenedor, desafortunadamente, existen algunas incompatibilidades.

Focal lenght: *Distancia focal* (de una lente).Distancia entre el eje óptico de la lente y el foco (o punto focal). Para una lente positiva (convergente), la distancia focal es positiva y se define como la distancia desde el eje central de la lente hasta donde un haz de luz colimado que atraviesa la lente se enfoca en un único punto. Para una lente negativa (divergente), la distancia focal es negativa y se define como la distancia que hay desde el eje central de la lente a un punto imaginario del cual parece emerger el haz de luz colimado que pasa a través de la lente.

FIFO: (FIRST-IN-FIRST-OUT) *Primero en entrar, primero en salir.* Dícese de las memorias de colas y de ciertos tipos de registros de desplazamiento.

Firewall: Cortafuegos. Sistemas de protección software y hardware, contra la intrusión de extraños vía Internet, en redes privadas de comunicaciones.

Field y frame: *Campo y cuadro.* En el sistema de TV utilizado en Europa, existen 50 campos y 15.625 líneas de exploración horizontal por segundo (625 líneas por 25 cuadros) para crear la imagen (en la cámara) o reconstruirla (en el receptor). Se utiliza exploración entrelazada para reducir el parpadeo de la imagen. Con este procedimiento, el sistema explora la mitad (312,5 líneas) de las líneas (línea por línea) durante una exploración vertical y seguidamente explora las otras líneas intercaladas con las primeras durante la siguiente exploración vertical desde la parte superior a la inferior de la imagen (las 312,5 líneas restantes). Son necesarias dos exploraciones verticales sucesivas para recorrer la totalidad de las líneas horizontales en que se divide la imagen (312,5 x 2 = 625 líneas) y la totalidad de una imagen; en otras palabras, puesto que se precisan dos exploraciones verticales para completar una imagen, en un segundo se tienen 25 imágenes completas. El término çampo"se refiere a la imagen parcial descrita durante un período de exploración vertical (1/50 segundo). Un çuadro.[es] la imagen completa producida por dos exploraciones verticales (1/25 segundo = 2 campos).

Frustum: Es la porción de un sólido, normalmente un cono o pirámide, que queda entre dos planos paralelos que cortan el sólido . En gráficos 3D es el espacio piramidal de proyección creado a partir de un sólido 3D y de la posición de dos planos de recorte, el plano cercano y el plano lejano, que delimitan un tronco de pirámide donde se sitúan los objetos que serán efectivamente visibles en la proyección. Los objetos que salgan fuera de estos planos

no se verán, y los que los intersecten serán recortados, dejando visible solamente la parte interior.

Getter: Cavidad circular de pequeño tamaño rellena de metales de rápida oxidación (como el bario) que sirve para evitar que los gases permanezcan en estado libre dentro de un tubo de aspiración. Estos metales consiguen la absorción del gas residual del tubo y bajan la presión del mismo y, se activan, con alto vacío, por calentamiento a una temperatura suficiente para provocar la difusión de la capa superficial «pasivada» en la masa. Esto produce la generación de una superficie metálica activa capaz de absorber las moléculas de gas presentes en el tubo.

GLUT: (OpenGL Utility Toolkit) API multiplataforma sencilla que provee una reducida funcionalidad para el manejo de ventanas e interacción por medio de teclado y ratón sin dependencia del sistema operativo, lo que permite una fácil migración de una plataforma a otra.

Jitter: *Inestabilidad.*Variaciones a corto plazo de las posiciones ideales en el tiempo de los instantes significativos de una señal digital. Variaciones del tiempo medio entre llegadas de paquetes.

Lent: Del inglés *lente*. Medio u objeto que concentra o hace diverger rayos de luz.

Line resolution :*Líneas de resolución* Técnicamente, el término se refiere a lineas verticales visualmente resolubles por altura de la imagen. Se miden contando el numero de lineas blancas y negras que se pueden distinguir en un area tan grande como el alto de la imagen. La intencion es hacer la medida independiente del formato. Las lineas de resolución horizontal se aplican tanto a la visualizacion en un televisor como a formatos de señales producidos por un lector de DVD.

Megapixel: *Mega píxel*. Unidad equivalente a 1.048.576 píxeles, usualmente utilizada para expresar la resolución de una imagen o de una cámara digital.

MPEG: Concepto que define un tipo de codec. Existen varias versiones, llamadas MPEG-1, MPEG-2, MPEG-4, ... pero es también un formato contenedor, a veces se denomina como MPEG Sistema. Existen varios tipos de MPEG: ES, PS, and TS Cuando se reproduce, por ejemplo, un vídeo MPEG de un DVD, el flujo MPEG está compuesto por varios flujos (llamados flujos elementales, ES): existe uno para el vídeo, uno para el audio, otro para subtítulos, y así sucesivamente que se juntan para formar un único flujo de programa (PS).

Multicast: Sin traducción directa. Las direcciones de envío múltiple (multicast) son direcciones de difusión como las de Ethernet, excepto que en lugar de incluir automáticamente a todos los nudos de la red, los únicos que reciben paquetes enviados a una dirección de envío múltiple son aquellos programados para escucharla. Esto es útil para aplicaciones como videoconferencia basada en Ethernet o audio para red, en los que sólo los interesados pueden escuchar. Están soportadas por casi todas las controladoras Ethernet (pero no todas). Cuando esta opción está activa, la interfaz recibe y envía paquetes de envío múltiple para su proceso. Esta opción corresponde al indicador ALLMUTI.

Multiplataform: Término inglés *multiplataforma* empleado en informática que designa la capacidad o características de poder funcionar o mantener una interoperabilidad de forma similar en diferentes sistemas operativos o plataformas.

NTSC: Sistema de televisión en color utilizado en USA, Canadá, Méjico y Japón donde NTSC M es el estándar de transmisión (M define el formato de campo y línea de 525/60 - con frecuencia el sistema se suele denominar simplemente NTSC). El ancho de banda en el sistema NTSC es de 4,2 Mhz para la señal de luminancia y de 1,3 y 0,4 Mhz para los canales de color I y Q.

Objetive: Término inglés *Objetivo* designa un conjunto de lentes convergentes y divergentes que forman parte de la óptica de una cámara tanto fotográfica como de vídeo. Su función es recibir los haces de luz procedentes del objeto y modificar su dirección hasta crear la imagen óptica, réplica luminosa del objeto. Esta imagen se imprimirá en el soporte sensible: sensor CCD o sensor CMOS en el caso de imagen digital, y película sensible en la fotografía tradicional.

OpenGL: (Open Graphics Library) Especificación estándar que define una API multi-lenguaje, multi-plataforma y escalable para escribir aplicaciones de gráficos 3D. Fue desarrollada originalmente por Silicon Graphics Incorporated (SGI). Podemos reseñar la inclusión de un lenguaje de shaders propio (GLSL) como estándar en la versión 2.0 de OpenGL presentada el 10 de agosto de 2004.

PAL: Fase alternada en cada línea (Phase Alternating Line). Sistema de codificación para televisión en color ampliamente utilizado en Europa y en todo el mundo, casi siempre con el sistema de 625/50 líneas/campo. Procede del sistema NTSC pero, al invertir la fase de la señal de referencia de color (burst) en líneas alternas (Fase alternada en cada línea) es capaz de corregir las variaciones de tono generadas por errores de fase durante el proceso de transmisión. El ancho de banda para el sistema PAL-I es de 5,5 Mhz para la luminancia, y 1,3 Mhz para cada señal diferencia de color, U y V.

Pixel (picture element): *Píxel o elemento de la imagen* es la menor unidad en la que se descompone una imagen digital, ya sea una fotografía, un fotograma de vídeo o un gráfico.

Quad buffer: Tecnología que permite a las tarjetas de video disponer de bufers independientes para la imagen izquierda y derecha, esto es, se usan en realidad cuatro bufers de imágenes. Mientras dos de ellos son mostrados a la vez, visualizando dos imágenes estereoscópicas, los otros dos bufers posibilitan la generación de las dos siguientes imágenes. Una vez las nuevas imágenes están listas, los bufers correspondientes se intercambian.

Rendering: Adaptación del inglés: *renderizado o interpretación.* Define un proceso de cálculo complejo desarrollado por un ordenador destinado a generar una imagen 3D o secuencia de imágenes 3D.

Reset: Del inglés *reponer o reiniciar*. Se conoce como reset a la puesta en condiciones iniciales de un sistema. Este puede ser mecánico, electrónico o de otro tipo. Normalmente se rea-

liza al conectar el mismo, aunque, habitualmente, existe un mecanismo, normalmente un pulsador, que sirve para realzar la puesta en condiciones iniciales manualmente

Resolution: *Resolución.* Cantidad de información gráfica que puede aparecer en una representación visual. En un dispositivo de representación en pantalla se indica por el número de líneas que pueden distinguirse visualmente. También se define la resolución de un sistema informático de gráficos por el número de líneas que se pueden representar en pantalla, o, de forma alternativa, por el número de puntos o píxeles (elementos de imagen) que pueden representarse en dirección vertical y horizontal medida en pixeles x pixeles .

Vertical Resolution: *Resolución vertical.* En sistemas gráficos de barrido, representa el número de líneas visualizadas por el monitor. En memorias de visualización, las distintas posiciones de la misma representan elementos de imagen a lo largo del eje vertical del visualizador. En sistemas de vídeo viene indicada por el número de líneas horizontales que pueden ser reproducidas y discernidas como elementos discretos en un monitor.

Shader: Tipo de lenguaje de programación destinado a programar el procesado de elementos (incorporación de nuevos efectos y propiedades aplicables en la definición de un objeto) de cualquier interfaz 3D o tarjeta gráfica (píxels, polígonos etc.).

Vision: *Visión.* Uno de los sentidos que consiste en la habilidad de detectar la luz y de interpretarla (ver). La visión es propia de los seres vivos teniendo éstos un sistema dedicado a ella llamado sistema visual o sistema óptico.

Computer Vision: *Visión por computador.* También conocida como visión artificial extiende la visión a las máquinas, o Visión técnica, es un subcampo de la inteligencia artificial cuyo propósito es programar un computador para que .[en]tiendaüna escena o las características de una imagen.

3 Introducción

3.1. Marco de Referencia

El proyecto que aquí se presenta bajo el título "Sistema estereoscópico para teleoperación asistida y supervisión de tareas Robotizadas", se engloba dentro de un proyecto de investigación del Instituto de Organización y Control de Sistemas Industriales (IOC), el proyecto TASTRI2 (figura 3.1).

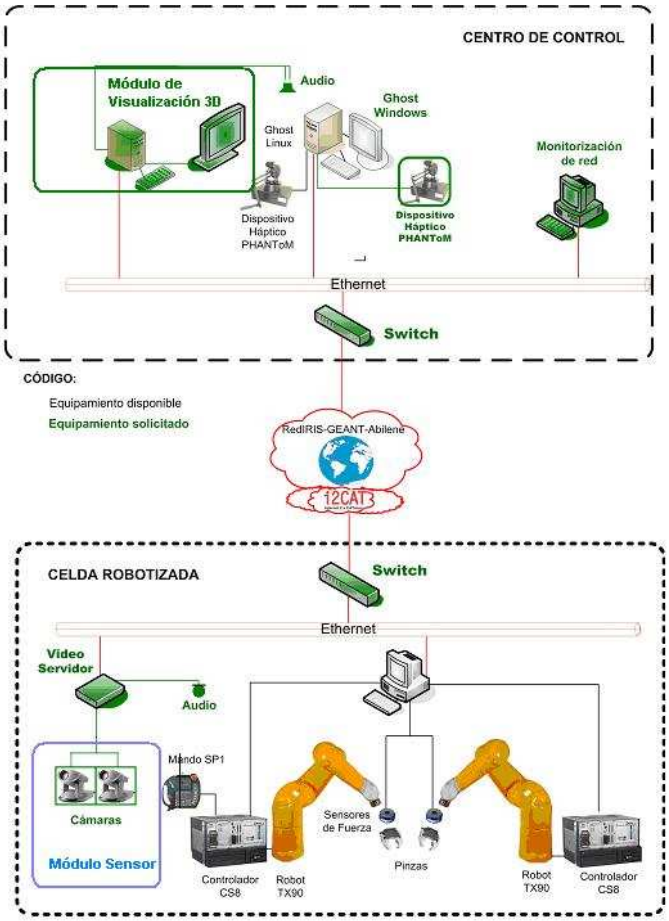

Figura 3.1 Proyecto TASTRI2

Este proyecto de investigación desarrollado con el apoyo y asesoramiento de diversas empresas industriales y, en particular con la Fundación I2CAT, tiene como finalidad agrupar entorno a la operación remota de sistemas multirobot en aplicaciones industriales, diversas disciplinas y herramientas como el posicionamiento relacional, la realimentación de fuerzas reales y virtuales, el guiado mediante reducción virtual de grados de libertad, la realidad aumentada y la comunicación a través de Internet2, que pueden dar una nueva dimensión no sólo en la utilización remota de sistemas de teleoperación sino también en ámbitos sociales, sanitarios, aeroespaciales y otros.

3.2. Objetivos

El proyecto que aquí se presenta pretende alcanzar uno de los objetivos parciales del proyecto TASTRI2, concretamente, se desea obtener un retorno sensorial mediante realidad aumentada y permite completar la información de un operador para el guiado de su operación dentro de una celda robotizada. Para ello, se implementan dos módulos distintos.

El primer módulo consta de un sistema sensor de imagen encargado del suministro de una parte esencial del retorno sensorial que, conjuntamente con la realimentación de la fuerza (que será recogido por otro módulo integrado también en el proyecto TASTRI2), se utilizará en la teleoperación y en la sintonización y supervisión remotas de la tarea.

El segundo módulo que contempla este proyecto se encarga de las tareas de visualización 3D que constituyen el soporte de la realimentación visual tridimensional de la celda, enriquecida con otros módulos del proyecto TASTRI2, por la simulación y realidad aumentada.

3.3. Alcance

En cuanto a equipo físico (hardware) se ha desarrollado un sistema binocular con cámaras de video conferencia de tres grados de libertad que pueden ser conectadas independientemente en dos puertos serie, o en cascada a un único puerto y, un sistema de representación de imágenes tridimensionales.

El software desarrollado en el proyecto ha consistido en una librería de control con arquitectura cliente-servidor, y, así mismo, una interfaz de usuario que permite la posibilidad de controlar cada una de las cámaras por separado o el control conjunto de las mismas; por otro lado, se ha desarrollado una interfaz que permite mostrar pares estereoscópicos, obtenidos o no por el sistema sensor de este proyecto, con el fin de ser mostrados en dispositivos de salida que permitan la visión tridimensional.

La descripción detallada de la arquitectura del sistema y un esquema conceptual del mismo se encuentran detallados en las secciones 6 y 7 correspondientemente.

3.4. Motivación

La comunicación multimedia en tiempo real, a través de paquetes de red, ha tenido gran atención en el último tiempo. Sus aplicaciones, como muetra la figura 3.2: simulacion de cirugía guiada remotamente, control de sistemas remotos, videoconferencias, aplicaciones de realidad virtual y realidad aumentada en sistemas colaborativos, entre otros muchos, han impulsado, por el lado de las telecomunicaciones, el desarrollo y estandarización de protocolos y tecnologías para transportar videos en tiempo real en redes IP.

Figura 3.2 Aplicaciones de la estereoscopía

Aunque hasta hace poco, la transmisión de vídeo a través de Internet tenía importantes limitaciones tecnológicas (ancho de banda pequeño, alta latencia y complejidad computacional) y, consecuentemente, la mayoría de sistemas de video tenían baja calidad y uso limitado, gracias a recientes avances en tecnología de redes y procesamiento de señales, se están eliminando rápidamente estos inconvenientes. Un ejemplo de ello, en lo que concierne a telecomunicaciones, es el nacimiento de Internet2, una red capaz de responder a estas exigencias haciendo uso de un ancho de banda capaz de llegar a los 40 Gbps en 2006.

Una motivación esencial para el desarrollo del proyecto tratado es cubrir las necesidades demandadas por este tipo de sistemas. Éstas sólo pueden conseguirse dotando del mayor realismo posible las comunicaciones de vídeo.

Muchos de los sistemas de transmisión de datos de vídeo, desarrollados hasta el momento, están limitados a imágenes monoscópicas, por lo que la percepción de tridimensionalidad resultante se pierde. La motivación se concreta en poder transmitir no sólo imágenes, sino indicadores de profundidad visual que de manera natural son perceptibles en el sistema de visión humano y hacerlo del modo más parecido posible.

Por todo ello, resulta de gran interés integrar, dentro de TASTRI2, dispositivos que doten de realismo las imágenes y permitan a un operador remoto llevar a cabo tareas teleoperadas como

si se hallase delante de los propios robots y estuviese manipulándolos. Para ello, debe ser posible la comunicación visual estereoscópica a través de paquetes de red y transmitir dos fuentes de vídeo (correspondientes a las obtenidas con los ojos) que terminarán por formar un sistema de visualización 3D y mostrarán correctamente dichas imágenes en un punto remoto gracias al uso de redes de comunicación, concretamente Internet, y con el tiempo Internet2. Una posible aplicación con retransmisión de imágenes y realidad aumentada se presenta en la figura .

Figura 3.3 Aplicaciones de Realidad Virtual y Realidad Aumentada

4 Antecedentes y Estado del Arte

4.1. Sistemas Estereoscópicos

Se comentan a continuación diferentes dispositivos que permitan la obtención de imágenes estereoscópicas.

Un sistema estereoscópico es un dispositivo que consigue que dos imágenes de dos vistas distintas, la correspondiente al ojo izquierdo y derecho, se muestren separadamente en el correspondiente ojo. Para conseguir dicho propósito, se han sugerido numerosas técnicas. Se comentan a continuación las más comunes.

4.1.1. Anaglifos

En este tipo de sistemas las imágenes derecha e izquierda se dibujan en diferentes colores: verdes-rojos, rojos-azules, o ámbar-azules (figura 4.1). El espectador, por su lado, es portador de un par de lentes con filtros de colores complementarios a las gafas, éstos cubren correspondientemente el ojo derecho e izquierdo, de modo que las imágenes de un color no serán vistas por el ojo portador del filtro del mismo color, y por ello, se consigue una correcta representación de las imágenes en cada ojo.

Este sistema, por su bajo coste, se emplea sobre todo en publicaciones, y en el cine. Sin embargo, presenta el problema de la alteración de los colores, por lo que este método no funciona correctamente con imágenes a color, así como una pérdida de luminosidad y cansancio visual después de un uso prolongado.

Figura 4.1 Sistema Anaglifo

4.1.2. Cascos estereoscópicos

Un casco estereoscópico es un dispositivo en forma de casco que el observador llevaría en la cabeza con un pequeño dispositivo de cristal líquido frente a cada ojo. En este dispositivo, la imagen izquierda y derecha se renderiza de modo distinto y separado en cada una de estas pequeñas pantallas . Suele incorporar, por regla general, un sensor que registra la orientación de la cabeza del usuario, y permite actualizar la imagen adecuadamente. El principal inconveniente es que tiende a ser poco confortable y sólo lo puede llevar un usuario cada vez. Por lo que se limita la compartición de experiencias entre diversos usuarios. Existen diversos tipos de cascos que se describen a continuación.

Cascos inmersivos

Estos cascos inmersivos también llamados HMD o Head-Mounted Display, (figura 4.2) aíslan al usuario de las imágenes del mundo real. Incorporan dos pequeñas pantallas con la óptica necesaria para permitir el enfoque. Cada pantalla está alineada con un ojo y recibe una señal de vídeo independiente.

Figura 4.2 Head-mounted display

Cascos HDC

Los cascos HDC (Head-Coupled Display), figura 4.3, son similares a los HMD, pero incorporan pantallas clásicas con cañones de electrones (Cathode Ray Tube, CRT) de alta resolución y pesadas; van montados sobre un soporte mecánico y, a su vez, contienen potenciómetros para monitorizar los movimientos del participante. Un ejemplo es el sistema BOOM (Binocular Omni-Oriented Monitor).

4.1.3. Sistemas basados en proyección

Son los más adecuados para el trabajo en grupo. En estos sistemas la imagen se proyecta sobre una o más pantallas (de monitor o, habitualmente, de proyección); éstas pueden adoptar diferentes disposiciones según su número y formato. Son instalaciones semi-inmersivas y en ellas el usuario puede ver su propio cuerpo y el entorno virtual.

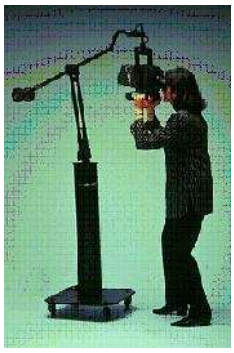

Figura 4.3 Sistema BOOM

Entre los diversos sistemas, se citan a continuación los más comunes.

Sistema CAVE

CAVE son las siglas de Computer-Animated Virtual Environment. Dicho sistema de visualización consiste en una habitación cúbica de 3x3x3 metros y de 4 a 6 paredes, las de cuatro son las más comunes. En el caso de una CAVE de 4 paredes, se proyectan imágenes tanto en las tres paredes laterales como en el suelo. Gracias al sistema de visión estereoscópica y a la interacción del usuario en medio de la proyección, éste tiene la sensación de que los objetos flotan dentro de la habitación, en lugar de distinguir las imágenes proyectadas individualmente en cada una de las pantallas (figura 4.4).

Figura 4.4 Sistema CAVE

Aunque el resultado obtenido en este tipo de sistemas es más que aceptable, su coste hace muy complejo su desarrollo e implantación.

Se presentan en el siguiente apartado otros dispositivos cuyo coste es, comparativamente, más asequible para un equipo de investigación.

Mesas estereoscópicas

Son sistemas que poseen una pantalla y en ella se proyectan las imágenes en estéreo mediante dos proyectores de características convencionales. Para ver imágenes en 3D con estos sistemas,

se utilizan filtros de doble polarización circular, y gracias ellos, la luz proyectada en la pantalla, luz polarizada, permite separar la imagen izquierda y derecha figura 4.5.

Estos sistemas no alteran los colores, pero sí una cierta pérdida de luminosidad y obliga al usuario a disponer de un espacio en el que colocar la pantalla y el sistema de soporte de proyectores. Su uso se extiende a proyección de cine 3D, monitores de ordenador con pantallas de polarización alternativa, etc, por ser un sistema económico con una calidad de imagen aceptable comparado con un sistema CAVE; sin embargo, no resulta tan inmersivo como este último.

Figura 4.5 Mesa estereoscópica

Gafas de obturación

Este otro tipo de dispositivo estéreo también está integrado por gafas para la visión estéreo. Dicho sistema usaría un esquema de visión multiplexado en el tiempo para mostrar las imágenes, y unas gafas con posibilidad de obturación "shuter glasses"(SH_G) de cristal líquido sincronizadas que restringen la visión de cada ojo a la imagen apropiada.

En dicho sistema las imágenes del canal izquierdo y derecho se presentan secuencialmente en un monitor CRT (de tubos de rayos catódicos) con una frecuencia sincronizada con las gafas de modo que el ojo izquierdo se descubre al proyectarse la vista izquierda y viceversa. La señal de sincronización usa generalmente un enlace infrarrojo, de modo que en el sistema puedan usarse simultáneamente diversas gafas como las que se muestran en la figura 4.6.

Figura 4.6 Gafas de obturación

Las ventajas de este sistema respecto a otros son: el uso simultáneo del sistema por más de un usuario, la no presencia de alteración de colores, y su luminosidad adecuada. Por otro lado, su coste es asequible y no obliga a disponer de un espacio reservado para su uso.

Después de estudiar las ventajas e inconvenientes de los diferentes dispositivos descritos anteriormente, dentro del proyecto TASTRI2, se opta por un sistema de estereovisión con uso de gafas de obturación, y salida de imagen hacia una pantalla o a un proyector 3D. Dicha solución se detallará más adelante como solución al módulo de visualización 3D de las tareas robotizadas de TASTRI2.

Pero para empezar a desarrollar un dispositivo de visualización estereoscópico, se precisa conocer primero los requerimientos de sistemas de este tipo. Como punto de partida, y por su familiaridad, se ha elegido el sistema de visión humano para establecer los requerimientos imprescindibles de estos sistemas.

5 Condicionantes técnicos del sistema

Podremos afirmar que la mayor parte de la información del entorno se recibe gracias al sentido de la vista y, de manera natural dicha información es estéreo: somos capaces de apreciar, a través de la visión binocular, las diferentes distancias y volúmenes en el entorno que nos rodea. Sin embargo, en el sistema visual se da la siguiente paradoja: por un lado las imágenes percibidas a través de los ojos y proyectadas en la retina son planas; por el otro, el mundo que nos rodea es tridimensional. El sistema de visión capta únicamente imágenes planas, pero gracias a la reconstrucción de las mismas en el cerebro a través de la asociación visual de diferentes fuentes (indicadores), junto con otros tipos de información, se consigue captar la profundidad de los objetos enfocados.

En este capítulo se comentarán en primer lugar, cuáles son los fundamentos del sistema de visión humano, y cómo el cerebro es capaz de distinguir objetos próximos a él de los que están más lejos. En segundo lugar, se describirán los conceptos fundamentales y los mecanismos de la visión y se definirá el concepto estereoscopía; se analizarán a continuación los sistemas que existen actualmente para conseguir generar una imagen estereoscópica para finalmente, describir el sistema elegido.

5.1. El sistema de visión humano

Para estudiar cómo el cerebro genera esas imágenes, se citan en primer lugar, aquellas partes de la anatomía del ojo cuya fisiología interviene en la captación de imágenes. Dichas partes representadas en la figura 5.1 son:

1. La córnea: Parte transparente de la capa externa del ojo.

2. Iris: Parte coloreada del ojo. Su función es regular la cantidad de luz que entra en su interior. Su parte central, por donde sí que puede pasar luz, se llama pupila.

3. Cristalino: Lente convexa por los dos lados, capaz de variar su curvatura, posibilitando el enfoque a diferentes distancias.

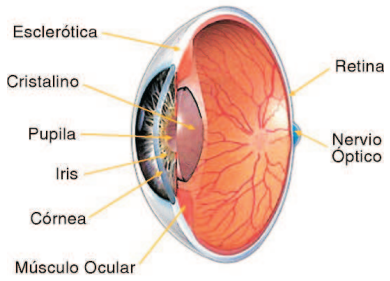

Figura 5.1 Anatomía del ojo

4. Retina: Capa posterior del ojo, donde se reúnen una gran cantidad de células foto-sensibles, especialmente en una región denominada fóvea.

5. Músculos oculares: Responsables de la movilidad del ojo.

Tratada de forma breve la anatomía del sistema de visión humano, se describe a continuación su fisiología, con el fin de comprender los mecanismos que permiten la visión tridimensional a partir de dos imágenes planas captadas por los ojos.

5.1.1. Fisiología de un sistema de visión

La fisiología del sistema de visión humano, al enfocar un objeto y generar su imagen, se podría resumir en los siguientes pasos:

1. En primer lugar, los músculos oculares dirigen el eje óptico hacia el objeto que deseamos enfocar.

2. El cristalino se acomoda para enfocar el objeto correctamente.

3. La luz procedente del objeto pasa a través del pupila y se proyecta en la retina. Concretamente, el objeto al que miramos se proyecta en la fóvea que es donde hay más células fotosensibles. Debido a la forma del cristalino, al pasar las imágenes, se invierten y se proyectan en la retina al revés.

4. Las células fotosensibles envían la información al cerebro a través del nervio óptico.

5. Por último, al recibir el cerebro las imágenes de ambos ojos, compone una sola dotada de profundidad.

Pero para establecer el paralelismo deseado con un sistema estéreo, es necesario estudiar más detalladamente los mecanismos que intervienen en el proceso previamente resumido.

5.1.2. El mecanismo de visión humano y la Estereoscopía

La estereoscopía es la estrategia más importante de un sistema de visión para obtener información de profundidad, tamaño y distancia a la que se encuentran unos objetos de otros.

Se basa esencialmente en la diferencia entre las imágenes que provienen de cada uno de los ojos, representada en la figura 5.2, que es debida a la separación inter-ocular, cuyo promedio, en un adulto, es de unos 6,5 cm.

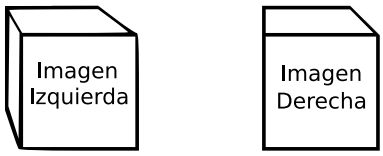

Figura 5.2 Diferencias de perspectiva de la visión de los ojos

Gracias a esta separación, el cerebro obtiene de una misma escena dos imágenes desde dos puntos de vista distintos, y ello se interpreta como una sensación de profundidad. A este proceso se lo denomina estereopsis. Y, a la capacidad de discernir detalles situados en planos diferentes y a una distancia mínima, agudeza estereoscópica.

En este mecanismo existen algunas peculiaridades y efectos derivados que vienen influídos por la distancia entre los ojos.

Por un lado, a mayor separación, mayor es la distancia a la que se aprecia el efecto de relieve. El efecto obtenido es empequeñecimiento de los objetos. Esta técnica se denomina hiperestereoscopía. Esto se aplica, por ejemplo, en los prismáticos, en ellos, mediante prismas, se consigue una separación interocular efectiva mayor que la habitual, con lo que se aprecia el relieve de objetos distantes que en condiciones normales no seríamos capaces de separar del entorno . También se aplica en fotografía aérea, donde se obtienen pares estereoscópicos con separaciones de metros, así es posible apreciar claramente el relieve del terreno que con la visión normal y desde gran altura sería imposible.

El efecto contrario se consigue con la hipoestereoscopía, es decir, con la reducción de la distancia interocular, imprescindible para obtener imágenes estereoscópicas de pequeños objetos (macrofotografías), o incluso obtenidas por medio de microscopios.

A parte de la estereopsis, intervienen otros mecanismos en el proceso de visión. Se detallan en la sección siguiente la convergencia, la acomodación y la fusión.

5.1.3. Otros mecanismos del sistema de visión humano

El mecanismo de convergencia aparece al querer observar objetos que se encuentran a diferentes distancias. Cuando se observan objetos lejanos, los músculos oculares restringen la geometría

de los ejes ópticos obligando a que se sitúen en paralelo. Sin embargo, al observar un objeto cercano, los músculos oculares giran los ojos para que los ejes ópticos se alineen sobre él, es decir, converjan.

Por otro lado, cuando se ha elegido el objeto que se pretende ver, la imagen deseada debe estar bien definida, entonces el sistema de visión recurre al mecanismo llamado acomodación (o enfoque), es el que permite ver el objeto de forma nítida.

Al proceso conjunto de estereopsis y enfoque, una vez llega al cerebro, se le llama fusión, es lo que realmente permite que se perciban las tres dimensiones de un objeto.

Cabe destacar que la fusión se realiza de modo diferente según la persona, es subjetiva. En distintos estudios empíricos, se ha constatado que no todo el mundo tiene la misma capacidad de fusionar un par de imágenes en una sola tridimensional. Hay una distancia límite a partir de la cual no somos capaces de apreciar la separación de planos, y varía de unas persona a otras. Así, la distancia límite a la que dejamos de percibir la sensación estereoscópica puede variar desde unos 60 metros hasta cientos de metros. Un factor que interviene de manera directa en esta capacidad de fusionar es la separación interocular, pero también cabe destacar, que el proceso de fusión de dos imágenes, mejora con la práctica a dicha exposición.

Una vez detallados los mecanismos que intervienen en la fisiología ocular, se estudian a continuación los indicadores y peculiaridades de los sistemas de visión por computador, en paralelismo con los humanos. Su finalidad es conocerlos en detalle para determinar cuáles debe tener el sistema diseñado.

5.2. La visión por computador

La visión por computador basada en la generación de pares estéreo, se usa para crear percepción de profundidad. Esta percepción puede ser empleada en ámbitos muy distintos, visualización científica (figura 5.3), entretenimiento de simuladores, juegos y apreciación de espacios arquitectónicos, etc.

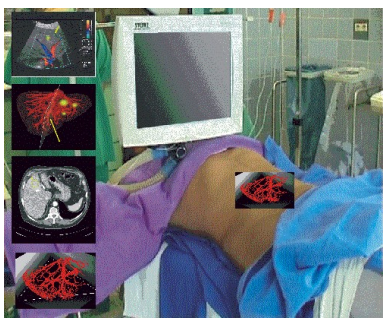

Figura 5.3 Aplicación de estereoscopía en visualización científica

5.2.1. Indicadores de profundidad

Cuando estamos viendo una imagen en un dispositivo de visualización convencional, una pantalla de ordenador, por ejemplo, los indicadores que tenemos para distinguir la distancia a la que se encuentran diversos objetos de la escena son, entre otros, la iluminación, las sombras, el tamaño de objetos conocidos o la interposición entre objetos; dichos conceptos se tratarán más adelante. Sin embargo, los objetos que vemos en pantalla carecen de volumen y, por ende, la imagen visualizada no resulta realista.

Esto se debe a que las imágenes sintetizadas en un dispositivo de visualización, al igual que lo que pasa con aquellas captadas por los ojos, son planas mientras que los objetos reales no lo son; por consiguiente se tratará de analizar el modo en que las imágenes sintetizadas son vistas por un observador como si poseyesen volumen. La respuesta se halla, como ya se ha anticipado, en una serie de elementos a los que llamamos indicadores de profundidad (depth cues).

Existen indicadores de profundidad utilizados en todos los sistemas de visión, humanos o computadores y que nos sirven para inferir información acerca de la distancia y profundidad de los objetos que nos rodean y ayudan a convencer al cerebro de que está viendo una imagen tridimensional.

Según una primera clasificación de los indicadores de profundidad podemos separarlos en dos grandes grupos: indicadores de profundidad monoculares e indicadores de profundidad binoculares.

Indicadores monoculares

Tradicionalmente las imágenes 2D crean su efecto de profundidad empleando indicaciones monoculares como los que se presentan a continuación:

1. Tamaño relativo: El cerebro ayuda a establecer un juicio acerca del tamaño de unos objetos por comparación con otros.
2. Perspectiva: Es una de las indicaciones de profundidad más importantes de los gráficos por ordenador. Se basa en que las líneas paralelas tienden a juntarse a medida que se alejan del observador.
3. Oclusión: Es la interposición entre objetos. Un objeto solapado a otro es percibido como más cercano.
4. Iluminación y sombras: El sombreado por efecto de la iluminación es una técnica básica, la luminancia de cada punto de una superficie nos proporciona información sobre la normal a la superficie. Igualmente los objetos pueden aparecer planos o curvados, lisos o rugosos, según el sombreado.

   Las sombras proyectadas también ofrecen información de profundidad. Esta técnica se utiliza ampliamente en publicidad, para simular que un objeto reposa sobre una superficie.

5. Perspectiva aérea: Los objetos con colores vivos y brillantes parecen más cercanos que objetos con colores apagados. Esto es debido a que la atmósfera absorbe parte de la luz que la atraviesa, especialmente las frecuencias cercanas al rojo. La niebla consiste en que el color de los objetos tiende a igualarse con el del fondo a medida que se alejan del observador.

6. Gradiente de textura: Un material con textura (por ejemplo una playa de piedras) proporciona una indicación de profundidad debido a que la textura se hace más aparente a medida que se acerca al observador.

7. Motion Paralax o movimiento relativo: Se basa en utilizar la velocidad relativa de los objetos para inferir información de profundidad. Los objetos lejanos parece que tengan un movimiento más lento que los del primer plano. Un ejemplo familiar son los postes telefónicos vistos desde un coche, que son atravesados más rápidamente que el paisaje de fondo.

Indicadores binoculares

Los indicadores de profundidad binoculares que no aparecen en las imágenes 2D y que son característicos de la visión tridimensional se muestran a continuación [10].

1. Acomodación y Convergencia: Como ya se ha descrito en el subapartado 5.1.1, son las tensiones, musculares en el caso del ojo humano, que se necesitan para cambiar la distancia focal de la lente del ojo con el fin de enfocar correctamente a una profundidad concreta en el caso de la acomodación y la tensión requerida para rotar cada ojo para que se oriente concretamente en el punto de enfoque en la convergencia.

2. Disparidad retinal: La disparidad retinal es la distancia entre puntos homólogos medida sobre la retina. Es debida a que los ojos están separados horizontalmente la distancia interocular. A saber, dos puntos en ambas retinas son homólogos si proceden del mismo punto del mundo real (o sintético). La disparidad retinal está provocada por el hecho de que cada ojo ve el mundo desde un punto de vista diferente.

3. Paralax: Es el mismo concepto que la disparidad retinal, pero medido sobre la imagen en un monitor o pantalla de proyección (en el caso de la visión plano-estereoscópica (figura 5.4), generada a partir de dos imágenes 2D). Este parallax, con el dispositivo adecuado, induce una disparidad en la retina, y a su vez produce la estereopsis.

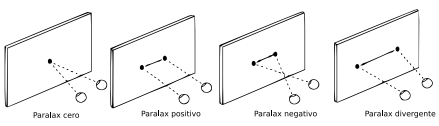

Figura 5.4 Tipos de Paralax

Por la importancia del paralax en sistemas de visión por computador dentro de la esteroscopía, se enumeran a continuación los cuatro tipos de parallax que existen:

Parallax cero: Representado en la figura 5.5, se produce al encontrarse el objeto en el plano de proyección, entonces la proyección en el plano focal coincide con los dos sensores de imagen y, por ende, se produce un paralax cero: los puntos homólogos ocupan exactamente la misma posición en cada imagen e induce una disparidad retinal nula. Por ello, en paralelismo con la visión humana, los puntos homólogos correspondientes al objeto donde convergen los ojos (objeto enfocado) tienen disparidad cero (son proyectados sobre la misma posición relativa de la retina), varios músculos mueven cada ojo de forma que la imagen que se está enfocando se sitúa sobre una posición en particular, la fóvea.

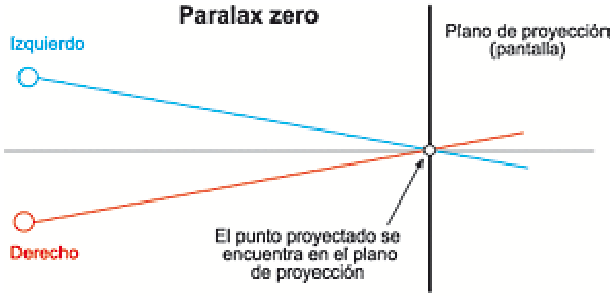

Figura 5.5 Paralax cero

Parallax positivo (no cruzado): La proyección está en el mismo lado que el respectivo ojo. Hay que notar que el máximo paralax positivo se da cuando el objeto se halla en el infinito; en ese punto el paralax horizontal es igual a la distancia interocular. Esto ocurre cuando los dos ejes ópticos son paralelos y se pretende ver en el mundo real los objetos lejanos. Todo valor positivo de parallax (entre 0 y la distancia interocular) produce imágenes que parecen estar detrás de la pantalla (figura 5.6).

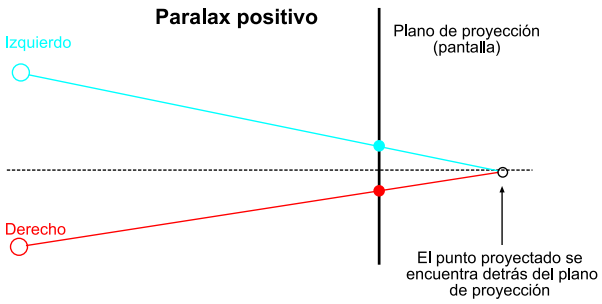

Figura 5.6 Paralax positivo

Parallax negativo (cruzado): Como se observa en la figura 5.7, los ejes ópticos son convergentes y se cruzan antes del plano de la pantalla. Objetos con parallax negativo parecen estar más cercanos que la pantalla, entre la pantalla y el observador. Si un objeto se halla

enfrente del plano de proyección, entonces la proyección para el ojo izquierdo está en la derecha y la proyección para el izquierdo en la izquierda. Cuando el objeto se mueve hacia un sitio más cercano al observador, el paralax horizontal negativo incrementa hasta el infinito.

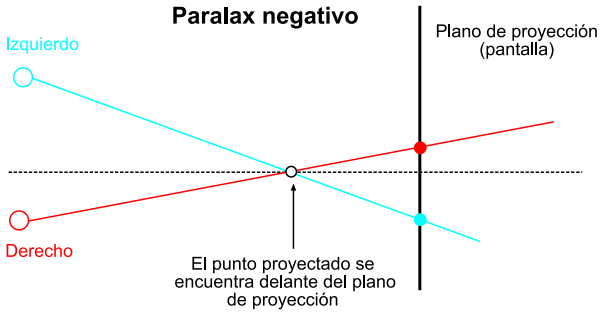

Figura 5.7 Paralax negativo

Parallax divergente: Si el parallax es mayor que la distancia interocular, los ejes oculares son divergentes y se produce fatiga ocular. Esta situación no se da en la visión normal y no es usada en estereoscopía.

La disparidad es considerada como el indicador de profundidad dominante, sin embargo, cuando otros indicadores se presentan incorrectamente se puede producir un detrimento del efecto 3D. Estos efectos entre los indicadores se presentan en el apartado siguiente.

Relaciones entre indicadores

Tradicionalmente, los indicadores monoculares y binoculares más usados para la representación de distancia, profundidad y estructura espacial han sido estudiados por separado asumiendo que la profundidad se procesa en módulos separados correspondientes a las diferentes fuentes de información tridimensional. Sin embargo, también es interesante resaltar cómo estos módulos, aunque independientes, pueden integrarse y relacionarse para producir una percepción 3-D coherente o, por el contrario, pueden entran en conflicto y hacer que se pierda toda sensación estéreo.

Existen diferentes modos en los que las fuentes de información pueden combinarse [16]:

1. **Promedio y sumatorio de indicadores:** Para indicadores sensoriales no intensos (dirección, orientación), la forma más efectiva de combinar indicadores es la combinación lineal por peso. Esto es, la profundidad independiente estimada para cada indicador o módulo de profundidad, se combina con los diferentes pesos asignados a cada indicador. Esta forma de interacción ha sido demostrada experimentalmente en varias ocasiones [14].

Para dimensiones sensoriales intensas (distancia) el sumatorio de indicadores nos proporcionará una percepción más detallada, si los criterios basados en indicadores individuales tienden a ser estimaciones a la baja. Con indicadores individuales concretos, el sumatorio nos conducirá a una amplia sobreestimación y tomar el término medio en lugar de la suma será lo más apropiado.

2. **Dominancia de indicadores:** Criterios basados en un solo indicador, donde el otro indicador es suprimido cuando entra en conflicto. Un ejemplo de esta situación en el contexto de un medio estereoscópico es el efecto borde de pantalla cuando una imagen estereoscópica se coloca ante una pantalla plana. La oclusión desde el borde de la pantalla dominará la percepción de profundidad y provocará que la imagen parezca curvarse en los extremos.

3. **Disociación de indicadores:** Cada indicador puede ser interpretado como surgido de un objeto distinto. Por ejemplo, cuando la separación espacial de señales del campo visual y auditivo de uno de los objetos excede de cierto ángulo, dos objetos pueden ser percibidos en lugar de uno, uno visual y otro auditivo. Un ejemplo conocido es percibir el vuelo de un avión a través de las diferentes localizaciones de sonido que origina.

4. **Reinterepretación de indicadores:** Uno de los indicadores puede ser interpretado de forma diferente después de combinarlo para hacerlo compatible con otro. Un ejemplo de este proceso es el efecto de profundidad cinética, i.e. donde la silueta de un objeto rotatorio, como un trozo doblado de cable, parece tridimensional incluso sin el indicador de disparidad, el cual se muestra adimensional cuando cesa el movimiento.

5. **Clarificación de indicadores:** Es un caso especial de reinterpretación de indicadores donde la señal del indicador puede ser ambigua (por ejemplo: cualquier objeto que esté frente o detrás del objeto fijado), y que requiere información suplementaria de otro indicador para ser interpretado. Un ejemplo podría ser una imagen borrosa, dota de información sobre la distancia de un objeto desde otro fijado sin indicar cual está más cerca. Otros indicadores como el de oclusión, tamaño conocido, o perspectiva lineal, pueden actuar como clarificadores.

Después de haber tratado a fondo el tema de indicadores de profundidad, se presentan en el siguiente apartado los requerimientos para la correcta representación de una imagen en un sistema de visión.

5.2.2. Resumen de requerimientos para el correcto renderizado de imágenes

Para representar un par estéreo de modo correcto se necesitan crear dos imágenes, una para cada ojo, de tal modo que, independientemente visualizadas, presenten una imagen aceptable en el cortex ocular. Si el par estéreo se crea con conflicto de indicadores, pueden ocurrir diversas cosas: un indicador puede resultar dominante y podría no ser el adecuado (la percepción de profundidad se reduciría), la imagen se volvería no confortable para ser vista, los pares estéreo no se fusionarían y el observador vería dos imágenes separadas.

Los pares estéreo crean una imagen tridimensional "virtual", si la disparidad binocular y la convergencia son correctas pero, si la acomodación es inconsistente, el resultado será que cada ojo mirará una imagen plana. Un sistema visual tolerará este conflicto de acomodación hasta cierto punto; la medida clásica está acotada con la máxima separación del dispositivo de 1/30 de la distancia del observador al dispositivo.

Hay varios métodos que permiten montar un dispositivo que pueda representar dos pares estéreo, muchos de ellos son estrictamente incorrectos dado que introducen paralax vertical causantes de disconfort y pérdida de la sensación de profundidad.

En el caso de dos cámaras con ejes cruzados, la proyección de dichas cámaras tienen una apertura simétrica; cada cámara apunta a un mismo punto. Las imágenes creadas usando dicho método seguirán apareciendo estereoscópicas pero el paralax vertical que introducen causa niveles de disconfort crecientes que incrementan desde el centro de proyección en el centro y son más importantes a medida que la apertura de la cámara aumenta.

Existe otro método estéreo que no introduce paralax vertical y además crea pares estéreo menos estresantes. Hay que notar que requiere un frustrum no simétrico de la cámara, que se soporta con algunos paquetes de renderizado, en particular OpenGL.

Los objetos que están en frente del plano de proyección aparecerán en frente de la pantalla del ordenador; los objetos que están detrás del plano de proyección aparecerán dentro de la pantalla.

El grado de efecto estéreo depende de ambas cosas: la distancia de la cámara al plano de proyección y la separación de las cámaras.

Separaciones grandes pueden ser complejas de resolver y se conocen, como se ha dicho anteriormente, como hyperestéreo. Una buena cifra aproximada de separación de las cámaras es 1/20 de la distancia del plano de proyección. Esto generalmente es la máxima separación para una visión confortable.

Otra restricción es asegurar que hay paralax negativo, es decir, el plano de proyección que está detrás de los objetos, no excede de la separación de los ojos. Una medida común del ángulo de paralax se define como:

$$\theta = 2\arctan(DX/2d) \tag{5.1}$$

donde DX es la separación horizontal de un punto proyectado entre los dos ojos y d es la distancia del ojo desde el plano de proyección.

Para facilitar el proceso de fusión de las imágenes, el valor absoluto de θ no debe exceder los 1,5 grados para todos los puntos de la escena. Notar que θ es positiva para puntos detrás de las escena y negativa para puntos en frente de la pantalla. Cabe decir que cuando el paralax

negativo es más difícil de fusionar las imágenes cuando los objetos cortan el límite del plano de proyección.

Así pues, el sistema esteroscópico que se diseñará deberá poder generar dos imágenes como las que aparecen en la figura 5.8, que corresponderán a la visión del ojo derecho e izquierdo correspondientemente.

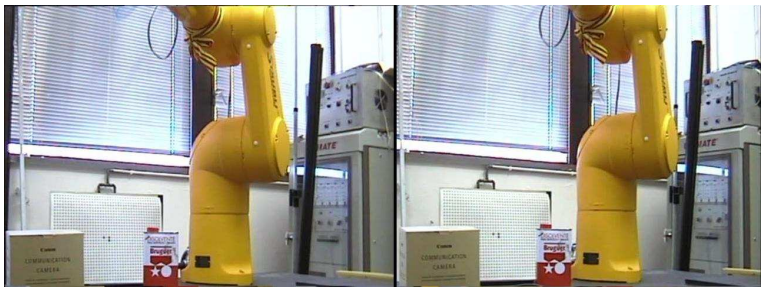

Figura 5.8 Par estereoscópico de una escena del laboratorio

Este par estéreo de imágenes, deberá tener consistencia de indicadores de profundidad. Deberá ser un método que permita representar dos pares estéreo, que no introduzca paralax vertical causante de disconfort y pérdida de sensación de profundidad, pero sí deberá conseguir que la imagen izquierda y derecha coincidan en todo: color, geometría y brillo, excepto en los valores del parallax, lo que permitirá al cerebro fusionar las imágenes de salida del sensor imagen y conseguir la sensación de profundidad deseada. De lo contrario, se producirá fatiga ocular.

Se detalla en la siguiente sección el diseño elegido y la implementación del mismo.

6 Diseño e Implementación del Módulo Sensor

Para el diseño del módulo sensor de este proyecto se ha considerado finalmente que este proceso estéreo sea simulado uniendo dos cámaras de video-conferencia idénticas con una separación inter-ocular adecuada, se codifican las señales de video y se transporta la información resultante a través de la red a uno o más receptores donde se decodifique dicha información y se muestre adecuadamente gracias al módulo de visualización.

6.1. Diseño del Módulo Sensor

Hay dos tipos de configuraciones de cámara que pueden ser usados para adquirir las imágenes con la solución propuesta:

- Configuración de cámaras con ejes paralelos
- Configuración de cámaras con ejes convergentes

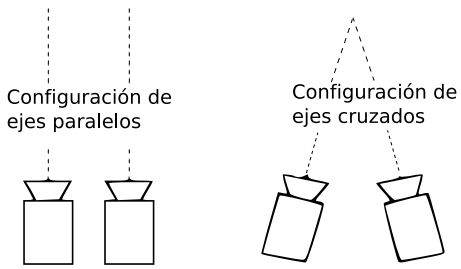

Figura 6.1 Configuraciones posibles de un sistema binocular

En ambas configuraciones, representadas en la figura 6.1, las cámaras deben estar alineadas verticalmente para no introducir paralax vertical que estropearía la sensación estereoscópica y, la separación interaxial debería ser de idealmente de 65 mm para conseguir dar una percepción de profundidad real. Sin embargo, otras separaciones serían también adecuadas para aplicaciones especiales como estereoscopía microscópica o mappings aéreos.

6.2. Diseño de las configuraciones posibles de las cámaras

6.2.1. Configuración de ejes paralelos

La configuración de ejes paralelos (figura 6.2) es aquélla en la que se alinean los ejes de las lentes de las cámaras de modo que los ejes ópticos de ambas funcionen de forma paralela. La convergencia de las imágenes se consigue al mover ligeramente las cámaras o con el traslado horizontal de las imágenes y con un recorte de las imágenes tratadas si es necesario [29].

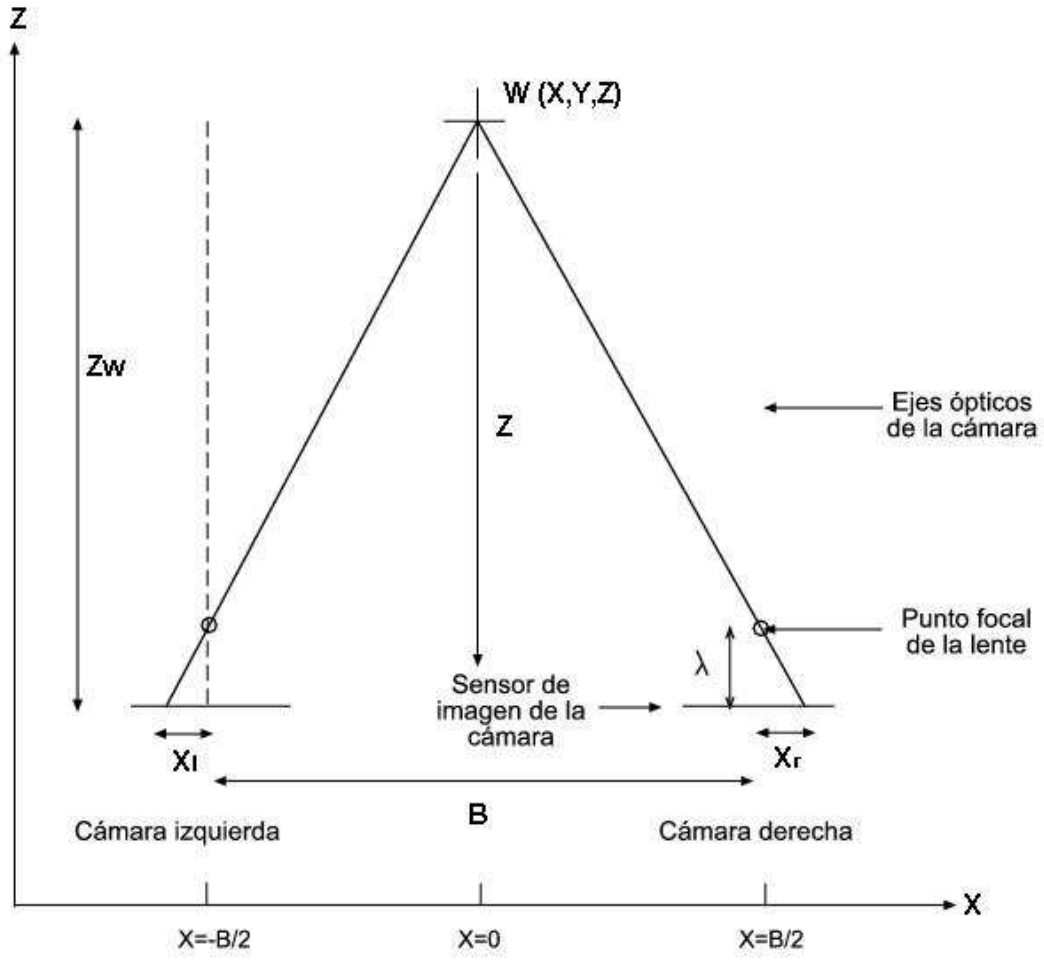

Figura 6.2 Disposición de dos cámaras en paralelo

Como se puede ver en la imagen anterior, un punto real, w, con coordenadas X, Y y Z, será proyectado hacia los sensores de imagen derecho e izquierdo y la disparidad horizontal será función de los siguientes factores: la separación base de la cámara, o línea de base, B, la distancia focal de las lentes de las cámaras, (λ), y la distancia de las cámaras al punto real, Zw.

En resumen, se puede ver el proceso de captura de las imagen de cada una de las cámaras como una traslación del eje X seguida de una transformación de la perspectiva. Las coordenadas de

proyección de la cámara (cámara izquierda: x_l, y_l, cámara derecha x_r, y_r) respecto al punto real w(X, Y, Z) son:

$$x_l(X,Z) = \lambda \cdot \frac{X + \frac{B}{2}}{\lambda - Z} \tag{6.1}$$

$$y_l(Y,Z) = \lambda \cdot \frac{Y}{\lambda - Z} \tag{6.2}$$

$$x_r(X,Z) = \lambda \cdot \frac{X - \frac{B}{2}}{\lambda - Z} \tag{6.3}$$

$$y_r(Y,Z) = \lambda \cdot \frac{Y}{\lambda - Z} \tag{6.4}$$

Como podemos determinar, comparando las ecuaciones 2 y 4, el punto real (X, Y, Z) se proyecta a la misma coordenada Y en las dos cámaras,y si las cámaras están correctamente alineadas no se produce desplazamiento vertical (o paralax vertical). La disparidad horizontal $d_{h,p}$, se obtiene sustrayendo x_l, la coordenada X de la proyección de la cámara izquierda, de la x_r, la coordenada X de la proyección de la cámara derecha:

$$d_{h,p} = x_r(X,Z) - x_l(X,Z) \tag{6.5}$$

De las relaciones anteriores, se puede obtener:

$$d_{h,p}(Z) = \lambda \cdot \frac{-B}{\lambda - Z} \tag{6.6}$$

De ello se deduce que la disparidad se incrementará con la separación de las cámaras, B, y con la distancia focal (λ). Hay que notar que para puntos en el infinito respecto al eje Z, la disparidad tiende a cero. Y, si se calibra el sistema apropiadamente, en la disposición paralela de cámaras no se produce disparidad vertical y, como consecuencia, tampoco distorsiones graves.

6.2.2. Configuración de ejes convergentes

En la configuración de ejes convergentes o cruzados, las cámaras se encuentran ligeramente rotadas la una respecto a la otra, de modo que los ejes ópticos de ambas cámaras interseccionan en un punto convergente y así toman importancia la separación de la base de la cámara (B) y el ángulo de convergencia (b) de los planos de cámara:

$$Z_{h,p} = \frac{B}{2 \cdot \tan(\beta)} \tag{6.7}$$

Esta disposición se muestra en la figura 6.3

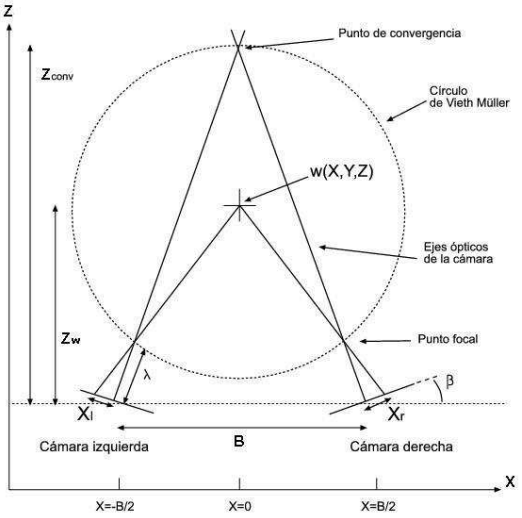

Figura 6.3 Disposición de dos cámaras con ejes convergentes

La proyección de un punto real (X, Y, Z) en el plano de imagen de la cámara es más compleja que en la disposición paralela de cámaras (es un caso especial de disposición convergente en el que beta equivale a cero). Como los ejes de la cámara ya no son paralelos a los ejes Z, se requiere una traslación horizontal (a lo largo del eje X) y una rotación (a lo largo del eje Y), seguida de una proyección de perspectiva. Para la cámara izquierda, la traslación horizontal es como sigue:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \xrightarrow{translación} \begin{bmatrix} X + \frac{B}{2} \\ Y \\ Z \end{bmatrix} \tag{6.8}$$

Una rotación del punto trasladado alrededor del eje Y de un ángulo positivo + β:

$$\begin{bmatrix} X + \frac{B}{2} \\ Y \\ Z \end{bmatrix} \xrightarrow{rotación} \begin{bmatrix} \cos(\beta)(X + \frac{B}{2}) - \sin(\beta)Z \\ Y \\ \sin(\beta)(X + \frac{B}{2}) + \cos(\beta)Z \end{bmatrix} \tag{6.9}$$

Finalmente, la proyección de perspectiva [13] , [36] , nos da las coordenadas x_l e y_l para la cámara izquierda:

$$x_l(X, Z) = \lambda \frac{\cos(\beta)(X + \frac{B}{2}) - \sin(\beta)Z}{\lambda - \sin(\beta)(X + \frac{B}{2}) - \cos(\beta)Z} \tag{6.10}$$

$$y_l(X, Y, Z) = \lambda \frac{Y}{\lambda - \sin(\beta)(X + \frac{B}{2}) - \cos(\beta)Z} \tag{6.11}$$

$$\tag{6.12}$$

De modo similar, las coordenadas del plano de la cámara para la cámara derecha se encuentran en:

$$x_r(X,Z) = \lambda \frac{\cos(\beta)(X + \frac{B}{2}) + \sin(\beta)Z}{\lambda + \sin(\beta)(X + \frac{B}{2}) - \cos(\beta)Z} \tag{6.13}$$

$$y_r(X,Y,Z) = \lambda \frac{Y}{\lambda + \sin(\beta)(X + \frac{B}{2}) - \cos(\beta)Z} \tag{6.14}$$

$$\tag{6.15}$$

Al obtener las coordenadas de la cámara izquierda y las coordenadas de la cámara derecha se puede determinar la disparidad horizontal para la disposición horizontal de cámaras.

Para la disposición convergente de cámaras existe también una componente de disparidad vertical, que puede ser calculada de forma separada. Esta componente vertical es igual a cero tanto en el plano definido como X=0 como en el plano Y=0; sin embargo, es más amplia en los extremos de la imagen. El paralax vertical provoca, como se ha dicho anteriormente, distorsión de la imagen que incrementa en los extremos, una curvatura del plano de profundidad que provoca que los objetos situados en las esquinas de la imagen estereoscópica parezcan estar más lejos que los objetos situados en el centro de la imagen.

El problema se ilustra en la figura 6.4:

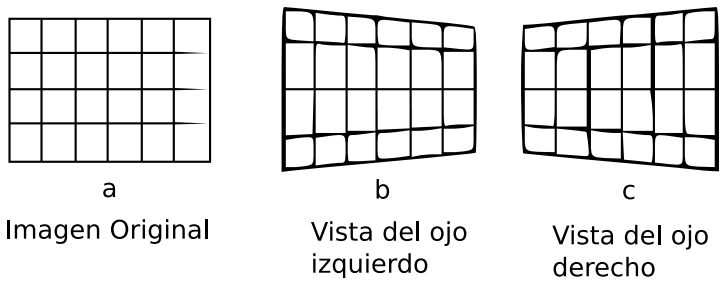

Figura 6.4 Efecto de distorsión en los extremos de la imagen

Con la formulación estereoscópica discutida es posible determinar de forma precisa la cantidad de disparidad tomada en el proceso de captación de una imagen, trasladando un objeto del espacio real (X, Y, Z) en coordenadas del plano de la cámara) (x_l, y_l) y (x_r, y_r). Cuando las imágenes estereoscópicas se presentan en un escenario, las coordenadas del plano de la cámara necesitan ser transformadas en coordenadas de pantalla. Esto se consigue multiplicando las coordenadas del plano de la cámara por el factor de magnificación de la pantalla M, que es la relación entre el ancho de pantalla horizontal y el ancho del sensor de cámara.

Los algoritmos de eliminación de esta distorsión se proponen en diferentes métodos, pero se evita con mayor facilidad al usar la configuración de cámaras paralelas. Aunque que la configuración de cámaras paralelas no da paralax vertical (siempre que las cámaras estén correctamente alineadas

verticalmente), en algunos casos se requiere una traslación horizontal de las imágenes resultantes. Debido a esta traslación, las imágenes no están perfectamente sobrepuestas. En este caso se tratarían las imágenes con un clipping de las mismas de modo que la parte común de la vista sea lo único que se proyecte. Dependiendo de cuánto estén trasladadas las imágenes, el plano de convergencia puede ser posicionado a diferentes profundidades.

6.2.3. Resumen del diseño elegido para el proyecto

Se ha diseñado un modelo de cámaras con ejes paralelos para iniciar tareas de telerobótica aumentada en lugar de la telerobótica usada en el laboratorio del IOC, dado que a priori presenta ventajas de control respecto a la misma:

1. Presenta invariancia en el tiempo de respuesta: Debido a que el operador recibe el feedback visual directamente de la simulación.

2. Una mejora del dispositivo: se introduce un dispositivo estéreo que mejora la visión de profundidad, el reconocimiento de objetos desconocidos en el espacio de trabajo,así como los gradientes y la orientación de los objetos.

3. El cálculo de trayectorias y simulaciones se puede realizar localmente aunque la escena sea remota.

4. Una reducción de los errores en la manipulación: Se reducen errores de colisión en la simulación y se prevén, a simple vista, configuraciones del robot no deseadas.

5. Control a alto nivel: El operador controla localmente la simulación en un ordenador local, desde él se envían instrucciones de alto nivel, como puntos de trayectoria que el robot deberá seguir.

Se presenta en el siguiente apartado el modelo de cámara que permitirá a largo plazo obtener las ventajas anteriormente presentadas.

Modelo de cámaras elegido

Se presenta en este apartado 6.2.3, una breve explicación sobre las características y criterios que llevaron a la elección de las cámaras Canon VC-C5 como entrada del módulo sensor de imágenes respecto a otras opciones posibles en el mercado.

En primer lugar, se buscaron modelos de cámaras que permitiesen la supervisión de la celda robotizada, para lo que sería necesario que tubiesen res grados de libertad. Por lo que se estableció la siguiente lista de opciones, todas ellas ideales para aplicaciones de videoconferencia, características necesarias para el tipo de aplicaciones de teleoperación del proyecto TASTRI2.

Lista	de	Modelos
Canon VC-C50i Infrared PTZ Camera		Canon VC-C50i Infrared PTZ Camera
Canon VC-C50i Infrared PTZ Camera		Canon VB-C50i Infrared PTZ Camera
Elmo PTC-100S PTZ Camera		Elmo PTC-201 CIP PTZ Camera
Elmo PTC-200C Inverted PTC Camera		Sony BRC-300 Robotic PTZ Camera
Sony BRC-H700 Robotic PTZ Camera		Sony EVI-D70 PTZ Camera
Sony EVI-D100 PTZ Camera		Sony SNC-RZ30N PTZ Network Camera

Cuadro 6.1 Lista de modelos de cámaras para la elección del HW del módulo sensor

En segundo lugar, por coste o porque se pretendía que fuesen controlables por software, se descartaron las Sony BRC, SNC y las Elmo.

En tercer lugar, por el conociemietno previo de la marca se seleccionaron las cámaras Canon, se descartaron el resto de modelos. Finalmente, la elección fue un par de cámaras Canon VC-C50i Infrared PTZ Camera, que además de presentar mejores prestaciones que su modelo anterior (VC-C4), están especialmente indicadas para videoconverencia, están dotadas de infrarojos, hecho que posibilita también la visión en ambientes con bajos niveles de iluminación o incluso, la visión nocturna (a 0 lux), tiene mayor rango de longitud focal (de 3,5 a 91mm) y poseen un circuito de reducción de ruidos, que permite obtener imágenes con gran nítidez.

El modelo elegido finalmente se presenta en la figura 6.5.

Figura 6.5 Imagen real del sistema elegido

Configuración y arquitectura elegida para el módulo sensor

Por sus limitaciones físicas, la cámara elegida del modelo Canon VC-C5i debe conectarse en serie con un PC. Gracias a ello, aunque el servidor de las cámaras esté cerca de ellas, por necesidad de estar conectado, via serie, con las dos cámaras, o en el caso de conexión en cascada de las

cámaras, conectado a una primera, y ésta a su vez, en conexión por cable con una segunda; pero al estar conectada a un pc servidor, cualquiera de las dos modalidades de conexión, permitirá que el cliente de las mismas pueda ser remoto.

Por este mismo motivo, se ha optado por una arquitectura cliente de comandos-servidor de movimiento de las cámaras, y se realiza la conexión entre ellos vía socket UDP-IP a través de las red, dado que la manipulación de la celda robótica y por ello su control pueda realizarse de forma remota.

Previamente cabría definir socket como la terminología que designa un concepto abstracto por el cual dos programas (posiblemente situados en computadoras distintas) pueden intercambiarse cualquier flujo de datos, generalmente de manera fiable y ordenada usando descriptores de fichero. Se ha elegido este tipo de conexión vía socket frente a la comúnmente usada TCP-IP por los siguientes factores:

En el caso de TCP (Transmission Control Protocol):

- Es un sistema orientado a conexión.
- En él se garantiza la transmisión de todos mensajes sin errores ni omisiones.
- Se garantiza que todo mensaje llegará a su destino en el orden transmitido.

Sin embargo, después de la recepción de cada uno de los mensajes, el servidor verifica la información y responde a cada cliente, según si el mensaje es o no correcto. En el primer caso, retorna una instrucción que da a entender que el mensaje es correcto, y procesa la petición. En el segundo, devuelve un mensaje y pide al cliente que vuelva a mandar otra petición por fallo o falta de elementos de la primera y se mantiene a la espera de recibir un nuevo mensaje.

Por otro lado UDP (User Datagram Protocol o Protocolo de datagrama de usuario).

- Está orientado a transporte.
- No hay control de flujos, no se garantiza la llegada del paquete a destino ni que no se produzcan errores.
- Su uso está especialmente indicado cuando no es posible realizar retransmisiones continuas por los estrictos requisitos de retardo, como cuando las aplicaciones son de Tiempo Real (TR).

Para nuestra aplicación, la pérdida de un paquete de información no es crítica dado que no se están realizando tareas de seguimiento de un objeto sino supervisando una escena remota que se teleopera en tiempo real; sin embargo, al ser una aplicación de tiempo real, no se quiere introducir en el sistema ningún tipo de retardo adicional que haga que la visión real de la escena

y lo que se muestre en el sistema tengan un desfase de tiempo, por eso el uso de sockets UDP, no sólo se justifica sino que se plantea como la solución más adecuada.

Además y al no tener confirmación de recepción de paquetes de información, al no esperar respuesta del servidor de su llegada, permite que el envío de nuevos mensajes pueda producirse a mayor velocidad. Esto es muy importante debido a una limitación del servidor actual de las cámaras, el envío simultáneo de mensajes a ambas cámaras. Limitación que se planteará como posible mejora del sistema actual en el apartado 9.

6.2.4. Configuración del los parámetros de las cámaras

Los sistemas com ópticos reales son de por sí, bastante complejos, pero en el caso que nos ocupa, existe un segundo factor a tener en cuenta, y es la característica de la cámaras de tener la lente motorizada, és decir, de zoom variable, lo que por un lado, aporta la ventaja de amplíar el rango de tamaños que pueden ser vistos y medidos. Sin embargo, por otro lado, deben tenerse en cuenta algunas consideraciones a la hora de emplear este tipo de lentes, ya que para poder mantener la relación de tamaños de los objetos que aparecen en la escena de las dos cámaras, deberán controlarse algunos parámetros que se presentan a continuación.

Por toda esta complejidad, para el desarrollo de este proyecto, se simplificará su comportamiento haciendo la hipótesis de que se trabaja con un sistema óptico de una sola lente modelada por la ecuación de Gauss de las lentes delgadas, aunque este modelo, difiera de los sistemas ópticos reales.

El tipo de sistema de visión, en el caso de una cámara digital, y con la hipótesis arriba planteada, es la que se muestra en la figura 6.6.

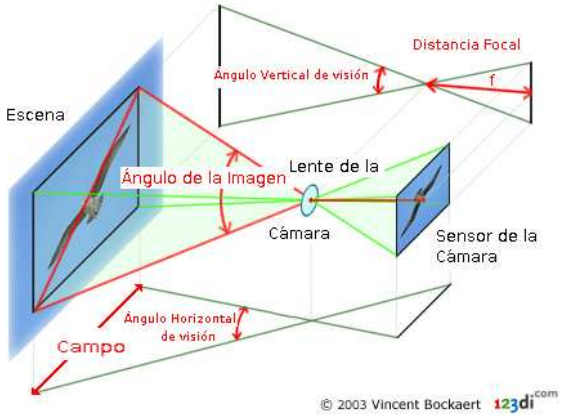

Figura 6.6 Proceso de formación de la imagen en una lente delgada

Cabe mencionar, que la ecuación de Gauss de las lentes delgadas a la que se hace referencia es la siguiente:

$$\frac{1}{F1} = \frac{1}{S1} + \frac{1}{S1'} \tag{6.16}$$

Donde F1 es la longitud focal, S1 la distancia de la lente al objeto real y S1' la distancia de la lente al plano sensor del CCD en el caso que nos ocupa.

Así, a causa de la lente y debido al movimiento de las cámaras dentro de la celda robotizada, se deberá cambiar adecuadamente la longitud focal de una de ellas, para que el tamaño de un mismo objeto que aparezca en ambas imágenes, sea idéntico cuando se encuentre a una distancia distinta de una cámara que de otra, cogiendo como ejes de referencia, la intersección de los ejes de giro de una de ellas.

La configuración de las cámaras en el sistema está recogida en la figura 6.7.

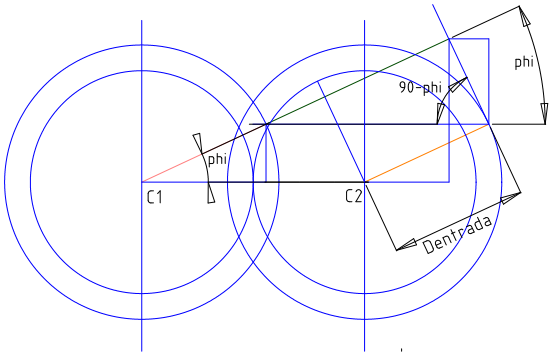

Figura 6.7 Configuración de las cámaras en el sistema:

El problema de la alteración de la longitud focal para obtener el mismo tamaño cuando el objeto a mostrar se encuentra a diferente distancia de la lente, se plantea en la figura 6.8.

Según este modelo de formación de imágemes, bajo la hipótesis de tener una lente delgada, se hallan, por triangulos semejantes las siguientes relaciones:

$$\frac{s1}{s1'} = \frac{y}{y'} \tag{6.17}$$

$$\frac{s2}{s2'} = \frac{y}{y'} \tag{6.18}$$

Siendo la ecuación de la lente, la planteada en la fórmula 6.16. se obtiene:

$$\frac{s1}{F1} = 1 + \frac{s1}{s1'} = 1 + \frac{y}{y'} \tag{6.19}$$

$$\frac{s2}{F2} = 1 + \frac{s2}{s2'} = 1 + \frac{y}{y'} \tag{6.20}$$

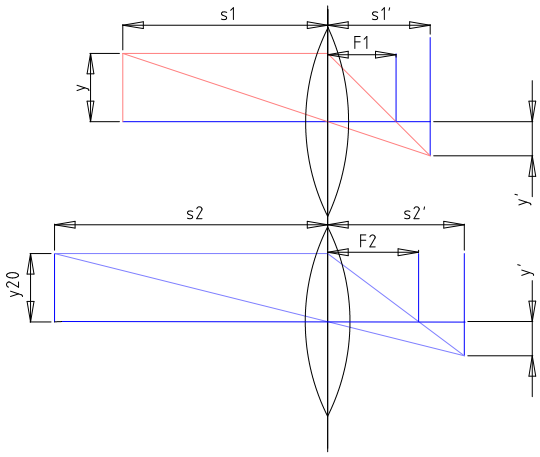

Figura 6.8 Proceso de formación de la imagen en una lente delgada

Igualando términos:

$$\frac{s1}{F1} = \frac{s2}{F2} \longrightarrow \frac{s2}{s1} = \frac{F2}{F1} \tag{6.21}$$

se consigue la ecuación que relaciona las diferentes distancias con las longitudes focales y que se utilizará para efectuar una corrección de los parámetros de la cámara que permita obtener el tamaño del objeto adecuado en ambas imágenes.

6.3. Implementación del Hardware

El sistema, tal y como se ha expuesto en la sección anterior, en su arquitectura, dispone de dos cámaras Canon VC-C5i que capturan la imagen del robot y su entorno. Las imágenes obtenidas corresponden a las vistas derecha e izquierda del ojo humano; se ha decidido colocarlas con sus ejes ópticos en paralelo a una distancia de 10 cm una respecto a la otra. Estas imágenes se presentarán al usuario y servirán pera la obtención del efecto de tridimensionalidad.

El dispositivo estereoscópico propuesto y que mejora los sistemas monoscópicos con la obtención, de manera interactiva, de información de una celda remota desde el propio sitio de trabajo, tiene un sistema de conexión entre cámaras y cámara-PC como el que se presenta en el esquema representado en la figura 6.9 presentada en las siguiente página.

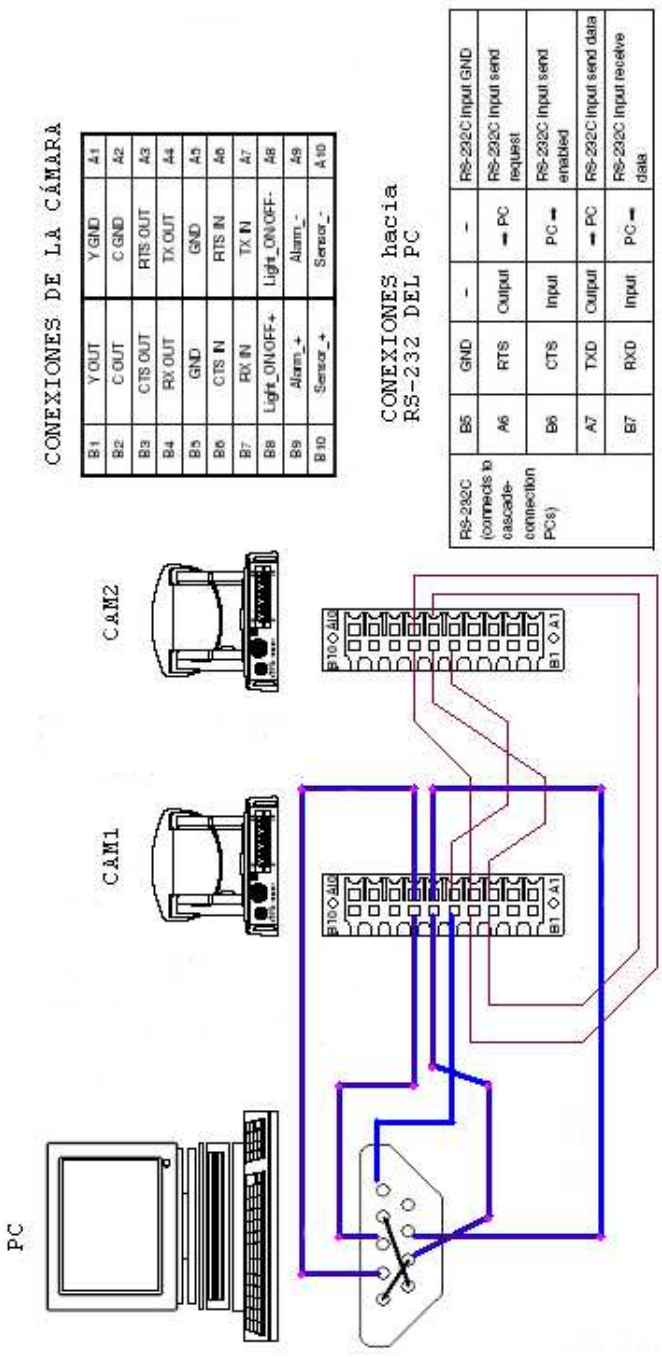

Figura 6.9 Conexiones del módulo sensor

Como se muestra y previamente se ha decidido en la fase de diseño, la conexión entre cámaras puede realizarse en cascada a través del conector (figura 6.9); las cámaras se conectan al puerto serie del PC en el que está en marcha el servidor de las mismas.

Por otro lado, como la conexión entre cliente y servidor se ha diseñado de modo que se produzca a través de sockets UDP, el cliente de la cámara podrá ser como se ha dicho,remoto. Pero, por motivos de protección de datos, se ha decidido que el cliente y el servidor se encuentren dentro de una área local que comprende todos los ordenadores de la UPC. Concretamente, aunque se espera que esta aplicación sea del todo remota y portable, se ha decidido que el ordenador Lira

sea el cliente, y sonar, el servidor de las cámaras. Los motivos: se hallan en el área local, sonar está cerca de las cámaras que deben conectarse vía serie con un PC, y Lira tiene la arquitectura de hardware necesaria para visualizar aplicaciones tridimensionales, gracias a la tarjeta que en él está instalada.

El sistema propuesto se describe a continuación en la figura 6.10.

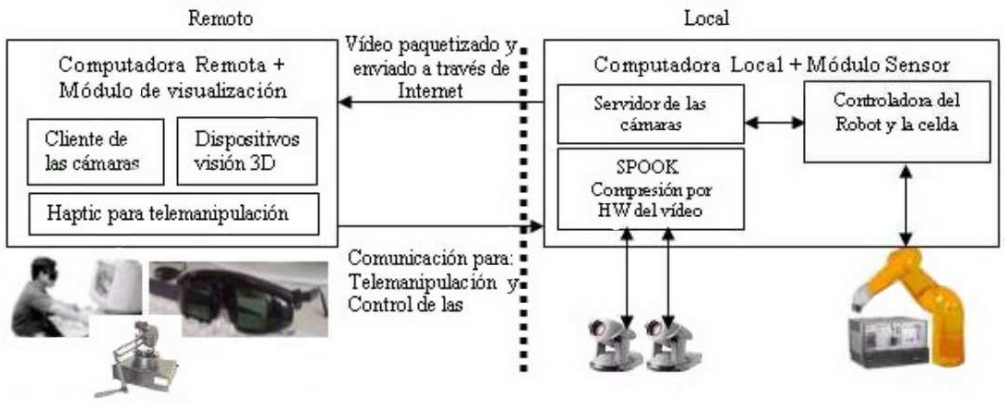

Figura 6.10 Diseño del esquema Cliente-Servidor del Módulo Sensor

De la parte del sistema de visión tridimensional propiamente dicho se hablará con mayor profundidad en las secciones Diseño e Implementación del Módulo de Visualización, por lo que sin más preámbulos, se trata a continuación la solución planteada desde el punto de vista de Software.

6.4. Implementación del Software

En este apartado se tratarán dos aspectos del software relacionados con el sensor de imagen, es decir, las dos cámaras. Por un lado, el servidor y el desarrollo de una librería de control que pone en funcionamiento y maneja las cámaras de forma directa y, por otro, en el cliente, la interfaz que permite el manejo de las cámaras por parte del usuario.

6.4.1. Servidor y Cliente del módulo sensor

Tal y como se ha comentado con anterioridad, estas cámaras se han programado según una arquitectura cliente-servidor remoto.

Para su desarrollo se ha partido de un antiguo servidor para cámaras de modelos canon más antiguos, concretamente de las cámaras Canon VC-C4 con las que ya se contaba en el laboratorio del IOC. Este antiguo servidor, desarrollado inicialmente por un equipo de investigación, cuya documentación y publicaciones en papers [23], [4] han servido de base y filosofía para el desarrollo del servidor que maneja el nuevo modelo de cámaras.

A partir de estos datos y con el manual de programadores descargable de la página oficial de Canon, [7] se ha desarrollado un servidor y éste se comunica con las dos cámaras vía puerto serie, y una librería programada en C++ a la que el cliente, a través de una interfaz de usuario, puede conectarse para controlar las mismas. Hay dos elementos esenciales que permiten el control de las cámaras;en primer lugar, un lenguaje comprensible por el usuario que puede ser transcrito al lenguaje máquina de las cámaras, y en segundo lugar un sistema de comunicación.

Por lo que al lenguaje se refiere, la librería de las cámaras, parte esencial del cliente, se comunica con el servidor de los dispositivos a través de mensajes de texto ASCII. El formato de estos mensajes está disponible en la documentación de este cliente en el anexo F.

Aunque no se detallará en este capítulo los pormenores de este lenguaje, cabe destacar que se compone de instrucciones y de símbolos de separación de mensajes [♯, @,];ello permite al usuario conversar con el servidor de control de los dispositivos y define el formato de los mensajes mandados entre el cliente y el servidor. El lenguaje no soporta explícitamente movimientos compuestos de las dos cámaras, sí permite realizar movimientos compuestos de una misma cámara: un mismo mensaje puede a su vez realizar una petición de rotación en dos de sus ejes.

Es importante saber que todos los mensajes contienen una marca de tiempo. Ésta será una marca de tiempo de la computadora, recogida en el momento en que el mensaje es construido por el cliente.

Por todo ello, se habla de una limitación de este lenguaje de las cámaras subsanado en trabajos futuros, y en el momento de realizar el movimiento de las dos cámaras, habrá un tiempo transitorio de movimiento de la cámara-1 en primera instancia, seguido del movimiento de la cámara-2; ambos, aunque seguidos, no se realizarán en el mismo instante de tiempo, y por ello

en este transitorio de este movimiento, la posición demandada de las cámaras y su posición real no se corresponderán

Por el lado de las comunicaciones, el cliente envía estos mensajes al servidor a través de la conexión UDP, y el servidor retorna una respuesta, a través de una comunicación con mensajes multicast, recogidos o no por parte del cliente . En resumen, el servidor, una vez se pone en marcha, queda permanentemente a la escucha del puerto UDP al que están conectados ambos, cliente y servidor, a la espera de comandos de texto en formato ASCII (peticiones de movimiento de la cámara).

Respecto a la comunicación con la segunda cámara, se ha habilitado para poderse realizar tanto con conexión vía puerto serie RS-232, como mandando los mensajes vía serie a la primera cámara (la más cercana al PC) para que los mensajes correspondientes a la segunda sean reenviados gracias a la conexión en cascada.

Interfaz de usuario para el control del módulo sensor

La interfaz de control de usuario se ha desarrollado gracias a las librerías de Qt desarrolladas por Trolltech cuyo enlace se encuentra en el anexo K (Enlaces). La funcion básica de esta interfaz de usuario es posibilitar el movimiento de las cámaras. Éstas por otro lado mandarán continuamente las imágenes que capturen mientras cliente y servidor estén en marcha. En primera instancia, estas imágenes pasarán al ordenador host conectado a estas cámaras; en segunda instancia, a través de la capturadora de vídeo, estas imágenes se comprimirán en formato MPEG4 para ser enviadas vía RTSP y finalmente descomprimidas y mostradas.

Esta interfaz de aplicación propiamente dicha se compone de diversos botones y barras de desplazamiento que pueden ser accionados, indicando qué movimientos realizarán las cámaras en cada momento, y cuál de ellas debe hacerlo; consta también de LCD programados para aportar información de la posición actual que según el cliente tienen las cámaras al mover o presionar alguno de los botones de la interfaz.

La interfaz diseñada tiene la forma presentada en la figura 6.11.

Aunque el diseño se ha realizado para ser intuitivo, merecería aclararse el uso de los siguientes botones y selectores de opciones.

Connect: Servirá para abrir la conexión UDP con el servidor a través del puerto elegido, el 5555.

Both: Indica que el movimiento pretendido por el usuario, se realice con las dos cámaras. Si el usuario sólo desea usar una de ellas, indicará previamente cuál de ellas y, a continuación, generará el movimiento deseado.

Disconnect: Se encargará del cierre de las conexiones UDP.

Exit: Permite salir de la aplicación.

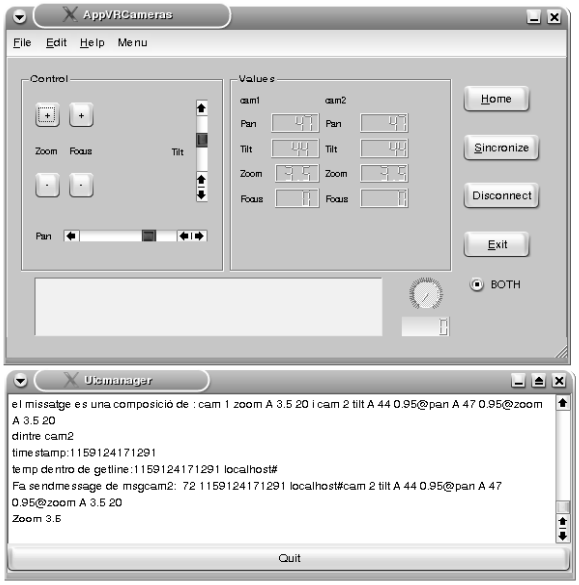

Figura 6.11 Apariencia de la interfaz de Qt que controla las cámaras

En el caso de Both, para que éste sea coherente y la imagen final sea estereoscópica, deberá resetearse la posición de las mismas mediante el botón home; éste inicializará a cero las posiciones de Pan (rotación derecha-izquierda) y Tilt (rotación hacia arriba y abajo), si previamente se han realizado movimientos por separado con cada una de las cámaras.

Cabe destacar que todas las librerías contenidas en este proyecto han sido seleccionadas por ser estandarizadas o, en su defecto, se han elegido aquéllas que, programadas por otras personas, tenían licencia pública de uso, con el fin de que todo lo desarrollado en este proyecto pueda ser usado, a su vez, por otras personas sin ánimo de lucro que puedan beneficiarse de ello. Además, el lenguaje usado tanto en la interfaz de Qt como en la documentación y comentarios de la librería de las cámaras LibVRCameras es el inglés; se pretende que el código pueda ser reutilizado y difundido para uso académico y docente.

7 Diseño e Implementación del Módulo de Visualización 3D

En este capítulo se presentan las posibles alternativas de diseño de un módulo de visualización estereoscópico y se destacan las diferencias más importantes entre ellos. Se ampliará a continuación la solución elegida para este proyecto y, finalmente, en la sección Implementación del hardware y software, se detallará a estos dos niveles la solución propuesta cuyo diseño, de manera conceptual, se ilustran en la figura 7.1:

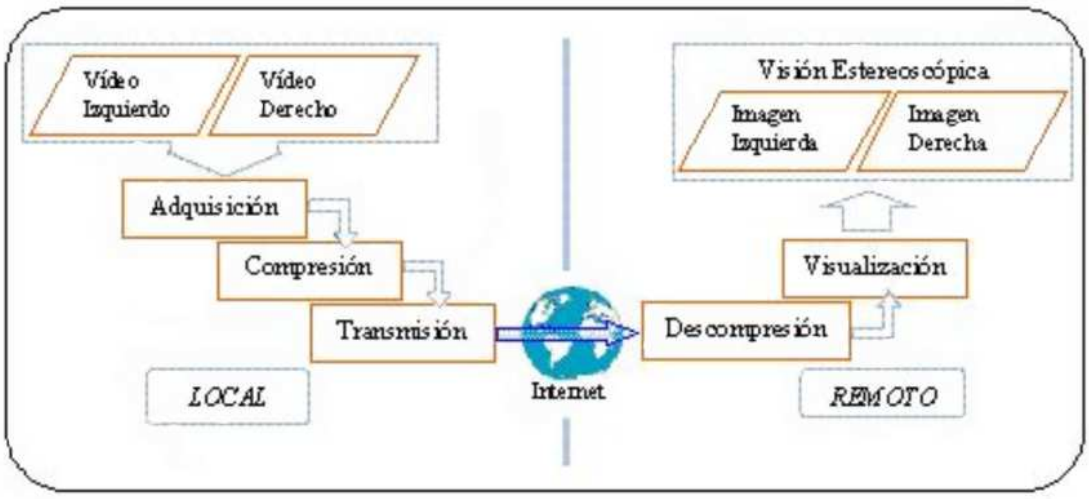

Figura 7.1 Diseño del esquema de transmisión

7.1. Diseño del Módulo de Visualización

7.1.1. Generación de las imágenes

La generación de las imágenes para ser mostradas por el módulo de visualización que aquí se presenta, se lleva a cabo en las cámaras descritas en el capítulo anterior mientras el usuario las mueve para supervisar las tareas de la celda robotizada.

Estas imágenes, pasan a través de dos cables de vídeo, a la tarjeta de adquisición montada en el PC servidor de las cámaras se muestra en la figura 7.2..

Figura 7.2 Tarjeta de adquisición de vídeo Adlink PCI-MPG24

Ésta tarjeta de adquisición de vídeo, modelo Adlink PCI-MPG24, recoge las imágenes recibidas por las cámaras a partir del momento en que se pone en el módulo de visualización. La imágenes se comprimen por hardware gracias a la tecnología de la tarjeta de adquisición de vídeo.

Compresión de frames

La compresión debe ser aplicada a las señales digitales de vídeo para usar el ancho de banda eficientemente, cuando se pretende mandar imágenes hacia un PC remoto. Numerosos esquemas de compresión se concebido para videos monoscópicos, y las técnicas y estándares aplicados se han retocado para poder comprimir vídeos estereoscópicos. Se reconoce que una compresión sustancial puede ser obtenida explotando la fuerte correlación que existe entre los canales de vídeo derecho e izquierdo de un par estereoscópico de vídeo.

Hay dos posibles codificaciones que explotan el hecho que las dos fuentes de vídeo sean parecidas: la codificación diferencial, y la codifiación por transformada. Aunque los pormenores de la codificación salen del ámbito de este proyecto se enumeran a continuación las principales características de estos dos métodos.

Por un lado, la codificación diferencial se usa cuando las diferencias entre muestras consecutivas son pequeñas, por lo que pueden transmitirse sólo las diferencias, aunque esta codificación puede ser con o sin pérdida de información. Por otro, en la codificación por transformada se aplica una transformada a la información a transmitir, por ejemplo, una transformada discreta del coseno, que permite pasar de una matriz de píxels a un a matriz de valores de frecuencias. Este método también puede darse con o sin pérdida de información

En este proyecto, por el parecido entre la disparidad del movimiento y la disparidad perspectiva, la compresión de vídeo puede beneficiarse de las técnicas predictivas de codificación diferencial desarrolladas para algoritmos de compresión temporal Inter-frame como el MPEG. La aproximación convencional para codificación de video, es codificar uno de los dos canales con un algoritmo de compresión monoscópica y codificar el segundo canal de modo diferencial respecto al primero, explotando las redundancias intercanal.

Aunque hay otros formatos de codificación de vídeo como el H.261 (estándar usado para videotelefonías y videoconferencia sobre RDSI sobre canales de 64kbps) o el H.263 (estándar para aplicaciones de vídeo sobre redes telefónicas e inalámbricas sobre canales de 28 a 56 kbps), se ha decidido usar una codificación en formato MPEG por ser un estándar ISO.

En el apartado siguiente se comentan por encima las diferencias entre los diferentes formatos MPEG y se decide cuál de ellos utilizar para la codificación del vídeo.

Codificación de vídeo MPEG

Dentro de la codificación de vídeo MPEG, existen diferentes codificaciones. Entre los más conocidos están los formatos: MPEG-1, MPEG-2, MPEG-7 y MPEG-4, aunque existen otros como el MPEG-2 Multiview Profile menos conocidos. Sin embargo, todos ellos, son todos estándares ISO.

Se indican a continuación principales características de los más populares. El primero, MPEG-1, se usaba originalmente para grabar Cdrom de audio/vídeo en calidad VHS a 1,5Mbps, más tarde se desarrolló el MPEG-2 que permitía transmisión y grabación de audio/vídeo en distintas calidades: en calidad baja, es compatible con MPEG-1 y alcanza los 4Mbps. En su calidad alta llega a tasas de 80 a 100 Mbps. Del formato MPEG-7 destaca que sólo contiene normas para describir contenidos multimedia y estructuras de información en otros estándares. Esos datos están orientados a ser usados por buscadores.

Finalmente, el formato MPEG-4, mejora a los sistemas anteriores por incluir normas para sistemas de poco ancho de banda y características de control de la señal de audio/vídeo. Está especialmente indicado para aplicaciones multimedia orientadas a Internet (de 5kbps a 50 Mbps), por lo que resulta el formato más indicado para el tipo de aplicaciones de este proyecto.

Una vez el vídeo procedente de las cámaras se ha codificado y el flujo de imágenes está comprimido en formato MPEG-4, se transmite a través de la red de Internet para ser posteriormente decodificado en un PC remoto y mostrado en dispositivos que permitan ver imágenes estereoscópicas. Los pormenores de la retransmisión de vídeo, se tratan a continuación.

7.1.2. Retransmisión de vídeo estereoscópico

Para la correcta retransmisión de vídeo a través de Internet, existen diferentes aspectos a tener en cuenta. Se presentan en los siguientes apartados aquellos necesarios para su diseño dentro del módulo de visualización aquí presentado.

Características de Internet

A la hora de transmitir vídeo sobre una red, en este caso Internet, debe conocerse el comportamiento de la misma y cómo puede afectar en la transmisión de datos mandados. Internet concretamente presenta tres características con un valor desconocido a priori y con carácter dinámico que afectan a los flujos de vídeo [5]:

1. Ancho de Banda variable.
2. Tiempo variable de transmisión.
3. Pérdida de datos.

El ancho de banda disponible entre dos puntos en Internet es desconocido y cambiante con el tiempo. Es decir, cuando el transmisor envía más volumen de datos de los que el receptor puede captar, se producirá congestión de la red y pérdidas de información provocando un deterioro en la calidad del vídeo observado; en el caso contrario, si el trasmisor es más lento que el receptor, la calidad del vídeo no será optima. Para paliar este efecto, el objetivo es adaptar la frecuencia de transmisión al ancho de banda disponible, ya que de esto depende la calidad en cuanto a resolución y actualización del vídeo.

Otro punto a tener en cuenta es el comportamiento variable que la transmisión punto a punto presenta, es decir, el tiempo que tardan en viajar dos paquetes de datos por la red es diferente, lo que puede causar un efecto de temblor en la imagen final debido a paquetes que llegan tarde.

Por último, las perdidas de datos dependen del sistema físico que caracteriza la la red. Ésta puede ser de par trenzado o inalámbrica. Normalmente, para contrarrestar el efecto causado y minimizar dichas pérdidas se utilizan técnicas de control de errores.

Protocolos para transmisión de flujos en Internet

Usualmente, la transmisión de paquetes de datos a través de Internet utiliza protocolos TCP/IP o UDP/IP. Como se ha mencionado anteriormente en el subapartado 6.2.3, el protocolo TCP (Transmisión Control Protocol) garantiza la entrega de información con retransmisiones y acuses de recibo por lo que están especialmente indicados para el control de flujo de información. UDP (User Datagram Protocol) es un protocolo que no garantiza la entrega de los datos más rápido por no enviar los datos de nuevo, por lo que se usan especialmente para la transmisión de datos.

Sin embargo, por sus características, estos protocolos limitan la transmisión de un vídeo, en el caso de un protocolo TCP por garantizar la entrega de información causan retrasos indeseables en estas aplicaciones; UDP mejora la velocidad de transmisión, pero no es fiable por no garantizar la llegada de los mismos. Por todo ello, se hace deseable el uso de un tipo de protocolo

especialmente dedicado a flujos de vídeo.

En este sentido, la IETF (Internet Engineering Task Force), comunidad internacional encargada de velar por el buen funcionamiento de Internet, ha especificado un conjunto de protocolos que definen el funcionamiento de la red para la entrega, control y descripción de flujo de medios sobre Internet. Para las descripciones ha creado la SDP (Session Description Protocol), para el control RTSP (Real-Time Streaming Protocol) o SIP (Session Initiation Protocol) y, para la entrega de medios, los protocolos RTP (Real-Time Transport Protocol) y RTCP (Real-Time Control Protocol). Las características principales de estos protocolos son:

RTP (diseñado para la transferencia de información): es un protocolo maleable, se puede configurar según el tipo de información que se quiere enviar, por ejemplo puede elegirse entre enviar video codificado en MPEG1 o en MPEG4 o audio en un tipo particular de codificación.

Sus características esenciales son:

- Se encapsula sobre UDP al estar asociado a aplicaciones de tiempo real.
- Tiene características especiales para el trabajo con sistemas de tiempo real como marcas de tiempo y números de secuencia (que permiten detectar pérdidas dentro de paquetes en un flujo de datos)

Pero como limitaciones cabe destacar que :

- No garantiza el envío de paquetes ni que el orden de llegada de los mismos corresponda con el orden establecido en el envío
- No proporciona mecanismos para el envío a tiempo de los paquetes
- No garantiza la calidad de servicio.

RTCP (diseñado para enviar mensajes de control): provee de una fuente periódica (cada cierto intervalo de tiempo se envían reportes de control) de realimentación al transmisor y receptores con datos relativos a la calidad de transmisión entre ellos.

Para obtener más información sobre el funcionamiento de este protocolo ver [28].

RTSP : se usa para establecer una sesión o conexión que controle el flujo de información entre transmisor y receptor con comandos para configurar, reproducir, pausar o grabar dichos flujos.

SIP : es usado para la transmisión de voz sobre IP y, aunque presenta características similares a RTSP, está provisto de funcionalidades adicionales.

SDP : es un protocolo usado para describir la sesión o conexión realizada, permite conocer si se transmite audio o video, los codecs usados, la velocidad y duración de la transmisión, protocolo a usar, etc.

Cliente y Servidor de Video

En el diseño de este módulo, para la transmisión y visualización de vídeo por Internet se ha decidido usar RTP por estar especialmente indicado para el trabajo con sistemas de tiempo real, junto a RTSP para el control del flujo de información a alto nivel, la descripción de protocolo de sesión (SDP) para configurar la conexión entre cliente y servidor de RTSP, y, finalmente, RTP/RTCP para ejecutar la información transmitida y conseguir su visualización en tiempo real.

La figura 7.3 muestra el esquema general utilizado.

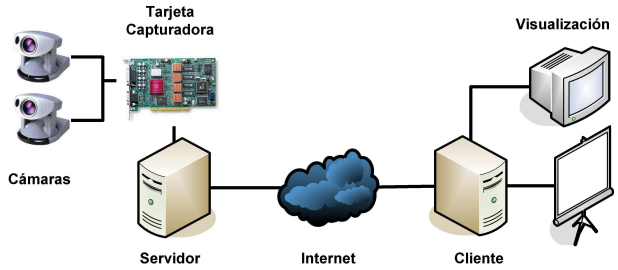

Figura 7.3 Esquema de Transmisión de video

El servidor se basa en una aplicación Linux llamada Spook que obtiene por un lado, un flujo de video o audio y lo transmite sobre redes IP (los detalles de esta aplicación se presentará en apartados posteriores); para éste proyecto, el flujo de vídeo a transmitir será el generado por la tarjeta capturadora de vídeo, con codificación por hardware MPEG4 y, por otro, utiliza RTP/RTCP para la transmisión de datos y RTSP junto a SDP para su control.

El cliente utiliza consecuentemente los mismos protocolos que el servidor realizando los siguientes procesos, conexión a un flujo de video, configuración de la conexión, recepción, decodificación y visualización de datos.

En la figura 7.4 se muestra el uso del protocolo RTSP entre el cliente y el servidor.

Cabe destacar que, al ser RTSP es un protocolo basado en mensajes, el servidor actúa respondiendo a las cinco posibles peticiones que se efectúen en el cliente.

Estas posibles peticiones se describirán al describir el desarrollo del software del módulo.

Una vez que el vídeo se ha transportado (comprimido en formato MPEG-4) y recogido por el sistema de visualización, es necesario separar cada uno de los vídeos que se han mandado y que corresponderían con los que viera el ojo izquierdo y derecho. Posteriormente, se decodifican los paquetes transportados y se obtienen del vídeo las distintas imágenes que se visualizarán. Este proceso, al realizarse por software, se detallará en el apartado 7.3.

7.1.3. Elección del método de visualización

El módulo de visualización necesario para formar parte del proyecto TASTRI2 debe ser un sistema no totalmente inmersivo y, por la limitación de espacio del laboratorio, no debe requerir un gran espacio para su implantación. Estas premisas permiten descartar gran parte de los sistemas de estereovisión presentados en el estado del arte de este proyecto.

De las alternativas posibles que cumplen con los requerimientos particulares de este proyecto, se han evaluado las ventajas e inconvenientes de los mismos, y se concluye que, de entre ellos, y por similitud con los mecanismos de visión humanos, debe desarrollarse un módulo de visualización con salida a un sistema de proyección y con gafas de obturación.

Para lograr la disparidad retinal y consecuentemente la estereopsis, es preciso que lleguen dos imágenes 2D al dispositivo de visualización, de forma que cada ojo perciba únicamente la imagen izquierda o derecha correspondientemente. Esto podría conseguirse de múltiples formas dependiendo del dispositivo visual utilizado.

Con la restricción de usar sistemas con salida a monitores o pantallas de proyección y gafas, actualmente existen tres tipos de sistemas de gafas estéreo: activos (multiplexado en tiempo), pasivos (multiplexado en espacio)y pseudoestereo.

Se detallan, a continuación, las características más relevantes de cada uno de ellos.

Estéreo activo

Para generar el estéreo activo con unas gafas, se precisa que cada una de las pequeñas pantallas muestre las imágenes derecha e izquierda alternadamente. Este hecho requiere el uso de frecuen-

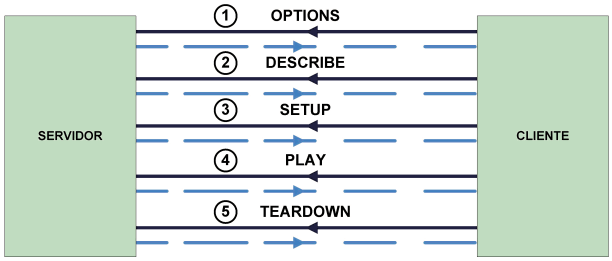

Figura 7.4 Esquema de Transmisión de vídeo

cias muy altas, de 120Hz, ya que las imágenes de salida deben dividirse entre los dos ojos y la frecuencia necesaria para cada uno es de alrededor de 60Hz. A su vez, las gafas que lleva el usuario del sistema, a parte de mostrar las imágenes, deben ser capaces de obturar los cristales alternadamente y de forma sincronizada con el sistema (ya sea mediante infrarrojos, cable de conexión, etc...) y, mientras uno de los ojos ve una imagen, el otro permanece opaco y viceversa.

El cerebro humano, al percibir tan rápidamente por cada ojo cada una de las imágenes, tiende a fusionarlas en una sola, con lo que se consigue obtener la sensación de estar viendo en tres dimensiones.

Existen algunos problemas con este tipo de sistema:

1. Si el sincronismo es por infrarrojos y no se poseen diversos aparatos colocados estratégicamente, se podría llegar a perder la "línea de visión"si algún objeto se interpusiera entre el emisor y las gafas.

2. Si la frecuencia de refresco no es muy alta, se percibe un parpadeo en las imágenes y puede ocasionar fatiga visual.

3. Las gafas, a parte de la obturación, generan las imágenes mediante pequeñas pantallas de LCD; acostumbran a ser pesadas y caras.

Cabe notar que, aunque antes la frecuencia de refresco era una gran limitación, hoy la mayor parte de fabricantes tienen máquinas y/o proyectores que admiten salida de vídeo a 120Hz y salida de sincronismo para las gafas estéreo activas.

Estéreo pasivo

Para la generación de estéreo pasivo se precisan: dos proyectores con filtros que polaricen la luz de forma inversa el uno respecto al otro (uno horaria y el otro anti-horaria) y que el usuario lleve unas gafas, en las que cada uno de los cristales tenga también un filtro polarizador.

Cada uno de los filtros de las gafas tendrá correspondencia con uno de los situados en los proyectores. De este modo sólo nos llegará una imagen a cada ojo, filtro dejará pasar únicamente la luz de una de las imágenes, la emitida por el proyector con el filtro correspondiente. A diferencia del método anterior, siempre existen dos imágenes simultáneas en la pantalla, ligeramente separadas, para crear un parallax artificial y que las imágenes se fusionen en al cerebro .

Como ventajas existentes del estéreo pasivo respecto al activo tenemos:

1. No requiere sincronismo entre las gafas y el sistema.

2. No existe la necesidad de doblar la frecuencia de refresco.

3. El problema de pérdida de "línea de visión"queda eliminado.
4. La fatiga ocular es menor.
5. Las gafas tienen el peso y ergonomía de unas gafas normales, y son muy económicas con respecto a las gafas de estéreo activo .

También existen inconvenientes como son:

1. Acostumbra a aparecer el fenómeno llamado "ghosting", visión parcial con un mismo ojo de la otra imagen debido a pequeños desajustes en la polarización.
2. Se debe duplicar el número de proyectores para generar una imagen en estéreo.

Modo de visualización pseudo estéreo

Este modo de visualización se asemeja en parte a los dos descritos anteriormente. Del modo activo se recoge el hecho de que las gafas son capaces de obturar los cristales alternadamente y de forma sincronizada con el sistema (ya sea mediante infrarrojos, cable de conexión, etc...) para que el ojo humano vea siempre una sola imagen. Del modo pasivo, que las imágenes no se proyectan directamente en las gafas; el coste y ergonomía de las mismas mejora respecto la alternativa del sistema activo.

Figura 7.5 Gafas obturadoras (shutter glasses)

Este modo se subdivide en otros dos, según la forma de mostrar las imágenes, según si la tarjeta de video tiene o no soporte estéreo del tipo Quad Buffer.

El primer caso, útil para gafas obturadoras (figura 7.5) cuando la tarjeta de video responde a modelos básicos, con soporte estéreo, no del tipo quad buffer, se basa en crear y enviar al monitor alternativamente cada imagen, coincidiendo con el refresco de pantalla y las gafas obturadoras sincronizan su parpadeo con este cambio.

El pseudocódigo utilizado para mostrar las imágenes es el que se transcribe a continuación:

```
Inicializar buffer de visualización
Si (la vista activa es la izquierda)
-->Mostrar escena para la vista izquierda
Sino
-->Mostrar la escena para la vista derecha
Finsi
Volcar el contenido en el buffer de visualización
Visualizar en el monitor el buffer de visualización
```

En el segundo caso, útil para gafas obturadoras cuando la tarjeta de video responde a modelos de gama alta, con soporte estéreo, del tipo quad buffer, se dispone de buffers independientes para la imagen izquierda y derecha. Así, es posible generar ambas imágenes simultáneamente. La tarjeta de vídeo es la que, de forma autónoma, envía alternativamente las imágenes al monitor de manera sincronizada con el refresco de pantalla. Las gafas obturadoras sincronizan su parpadeo con este mismo cambio.

El pseudocódigo utilizado es el siguiente:

```
Inicializar buffer de visualización
Tomar escena para la vista izquierda
Volcar el contenido en el buffer de visualización vista izquierda
Tomar la escena para la vista derecha
Volcar el contenido en el buffer de visualización vista derecha
Visualizar en el dispositivo los buffers de visualización
```

Finalmente, mencionar que el sistema elegido, por sus ventajas y bajo coste, es del tipo pseudoestéreo con tarjeta de video con soporte estéreo del tipo QUAD BUFFER modelo NVIDIA FX1100. Las imágenes se mostrarán a través de de una pantalla CRT o un proyector que estarán sincronizados mediante infrarojos con unas gafas obturadoras.

7.1.4. Sincronización del sistema de visión

Como se ha mencionado anteriormente, los dos flujos de vídeo vía RTP son transportados hasta el receptor, que debe ser capaz de sincronizar los flujos de video izquierdo y derecho a la hora de reproducir las imágenes. Esto es extremadamente importante cuando los objetos de la celda de robótica, pueden moverse en la escena, ya que serían percibidos valores falsos de páralax como resultado del desplazamiento espacial de los objetos en las dos vistas.

Gracias a usar un protocolo RTP, las marcas de tiempo de los dos flujos de vídeo, derivadas del mismo reloj, representan los instantes de muestreo de las imágenes de éste y pueden, por ello,

usarse para la sincronización. Esto relativamente sencillo si, como es el caso, los dos streams se originan en el mismo PC.

De otro modo el receptor se vería obligado a relacionar las marcas de tiempo RTP de los flujos de vídeo con sus marcas de tiempo NTP protocolo de marcas de tiempo de internet, correspondientes al informe de envío de paquetes proporcionado por RTCP y, consecuentemente, las marcas de tiempo de red (NTP timestamps) de las fuentes de transmisión de datos se podrían sincronizar.[20]

7.2. Desarrollo del Hardware

A medida que se han explicado las diferentes técnicas de visualización de imágenes tridimensionales y se diseñado el sistema propuesto para en este proyecto, han ido surgiendo los elementos necesarios para llevarlo a cabo. En este apartado se detallan cuáles se han elegido y con qué criterios se escogieron, así como sus principales características y cómo llevan a cabo sus funciones.

7.2.1. Elección de la tecnología del monitor

Existen en el mercado diferentes tipos de monitores agrupados en tres familias: de plasma, de tubo de rayos catódicos y los LCD. Cada una de éstas familias aplica un tipo de tecnología de proyección completamente distinta. Por lo que, para la elección de la tecnología del monitor integrado en este proyecto, se barajaron diferentes opciones y se eligió aquél con características más compatibles al tipo de aplicaciones de realidad aumentada pensadas para TASTRI2.

Cada una de éstas tecnologías usadas aporta características diferenciales a sus respectivos monitores.

Por un lado, las pantallas de Plasma utilizan fósforos excitados con gases nobles para mostrar píxeles y dotarlos de color. Su precio suele ser más elevado, pero la calidad también. En concreto el plasma ofrece mayor ángulo de visión que una pantalla LCD, mejor contraste y más realismo en los colores mostrados.

Por otro lado, las pantallas con tecnología LCD utilizan moléculas de cristal líquido colocadas entre diferentes capas. Estas moléculas pueden ser polarizadas y rotadas según si se quiere mostrar un color u otro. Cabe notar que cuando estas pantallas usan transistores TFT, se habla de TFT LCDs, que son los modelos más extendidos.

Finalmente, los CRT, o pantallas de tubos de rayos catódicos utilizan flujos de electrones a alta velocidad procedentes del un cátodo. Esta gran velocidad se debe a la alta tensión del ánodo. Los electrones son concentrados magnéticamente gracias a una bobina para obtener un rayo fino. La densidad del rayo puede ser controlada por una rejilla. Finalmente, éste es desviado magnéticamente por las bobinas llegando al ánodo cubierto de un material con tres tipos de

fósforo diferentes, los cuales emiten un color rojo, verde o amarillo. Cuando incide un haz de electrones golpeando ésta superficie, se emite luz.

Descritas brevemente las características principales de esos tres tipos de tecnologías, se presentan a continuación las ventajas e inconvenientes que aportan y que han tenido como consecuencia la elección de un CRT como dispositivo de salida.

Las principales ventajas de los LCDs en comparación con las pantallas de plasma, además del reducido tamaño, son el ahorro de energía que suponen los LCDs y, la mayor definición de las pantallas a igualdad de tamaño, (actualmente las pantallas LCD ofrecen mejor definición que las pantallas de plasma), aunque unas y otras pertenece a mercado o categorías diferentes.

Aunque por aportar mayor resolución podría pensarse que la tecnología de los LCD sería la opción más conveniente, existen importantes desventajas en el uso de monitores con ésta tecnología. En primer lugar, los LCD, en comparación con los CRTs, no pueden formar imágenes de resolución múltiple; sólo pueden producir imágenes claras en su resolución nativa o en una fracción menor de la misma. En segundo lugar, el ratio de contraste para los LCD es menor que la de los CRT y, finalmente, por su mayor tiempo de respuesta los LCDs pueden mostrar imágenes secundarias mezcladas con las nuevas imágenes cuando las imágenes cambian rápidamente, lo que produciría un efecto de ghosting indeseable en aplicaciones de estereoscopía.

Así pues, debido al tamaño de pantalla deseado, al coste de los monitores y al tipo de aplicaciones que en el proyecto TASTRI2 se pretenden desarrollar, se ha decidido usar, como se avanzó con anterioridad, un monitor con tecnología CRT como dispositivo de salida, ya que a parte de permitir una gran resolución, permite la aplicación de varias técnicas de estereoscopía. Desde la visualización de anaglifos, por ejemplo, con el uso unas lentes con filtros de colores; a la generación de imágenes entrelazadas para estéreo activo o pseudo-estéreo.

Aunque se haya elegido un monitor CRT para este proyecto, en un futuro, la opción a tener en cuenta sería el uso de SEDs (Surface-conduction Electron-emitter Display), una tecnología relativamente nueva y similar a la de tubo de rayos catódicos, que permite fabricar televisores tan planos como los de plasma o los LCD, por lo que mejorarían la ergonomía de las pantallas CRT. Además, consiguen una calidad de imagen mayor con imágenes en movimiento y, su consumo de energía es comparativamente más bajo (un tercio menor de la energía necesaria en un televisor plano de plasma, y dos tercios menor que la de un monitor LCD).

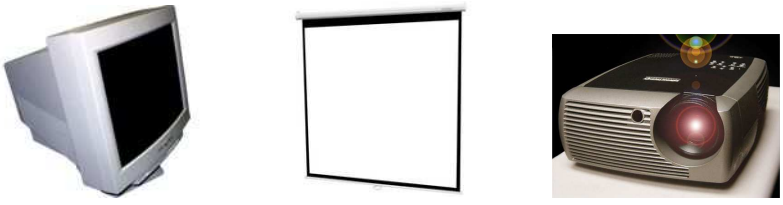

Figura 7.6 Dispositivos de salida de imagen elegidos para el módulo de Visualización

Lista de Modelos	
Cybertronics Universal Cybervision	e-LET'S Beautyful 3D
VNM Virtual-i	ELSA 3D Revelator IR
eDimensional 3-D Glasses	

Cuadro 7.1 Lista de modelos de cámaras para la elección del HW del módulo sensor

Como se aprecia en la figura 7.6, además del monitor CRT de Silicon Graphics, se ha añadido al sistema presentado en este proyecto, la posibilidad de usar también un proyector In Focus^{TM}DepthQ^{TM}3D como dispositivo de salida de imágenes, ya que este proyector, permite que las imágenes puedan ser proyectadas en una pantalla, capaz de adoptar diferentes formas. Podemos tener pantallas planas, esféricas, etc.

Además, el uso de un sistema de proyección, al emitir un haz de luz previamente generado por un dispositivo en el que está la imagen a proyectar por lo que tratando el haz de luz que sale del proyector antes de que la imagen se proyecte en la pantalla, proporciona a un sistema estereoscópico, la posibilidad del uso de la polarización como técnica estereoscópica.

Así pues, en los proyectores podremos usar las mismas técnicas estereoscópicas que en un CRT, con la ventaja de que además s podrían usar técnicas de polarización.

7.2.2. Elección de las gafas de obturación

De entre las opciones que ofrece el mercado de gafas de obturación para aplicaciones de estereoscopía, se eligieron en primer lugar, los modelos que cumpliesen con los siguientes requisitos: el controlador debía ser de tipo VGA por temas de compatibilidades con el hardware ya comprado para el proyecto, y las gafas debían ser no cableadas, con sincronización por infrarojos (wireless infrared shutter glasses), por ergonomía del usuario en las aplicaciones de supervisión de tareas, en las que éste puede estar observando la celda de trabajo y controlando diferentes parámetros de los robots que maneja, en cuyo caso puede necesitar mayor movilidad en su espacio de trabajo.

La lista de modelos obtenida después de hacer esta primera selección se presenta a continuación:

Finalmente, por temas de coste y conocimiento de marca, se eligieron las gafas de eDimensional, Cristal Eyes wireless (figura 7.7) se conectan mediante un cable a la salida de vídeo VGA, que junto a un mini emisor de infrarojos de largo alcance, que permien que el usuario pueda moverse en un radio de 3m. Los emisores de infrarojos a su vez, se encargarán de la sincronización con las con las gafas, necesarias para pasar, en cada una de las lentes, de las imágenes correspondientes al vídeo a la obturación de las lentes.

Figura 7.7 Gafas obturadoras del módulo de Visualización

7.2.3. Elección de la tarjeta de vídeo

Los criterios seguidos para la elección de la tarjeta de vídeo con soporte estéreo instalada, fueron los siguientes.

En primer lugar, se evaluó de entre las diferentes empresas de chipset de tarjetas de vídeo del mercado(ATI, NVIDIA), aquellas que para prestaciones 3D similares, tuviesen drivers para todos los sistemas operativos usados en el IOC, y en especial, para el sistema operativo Linux. Ante estos criterios, y respecto a las tarjeta en sí, existen diferencias entre ATI y nVidia, aunque ambas implementan OpenGL, ATI da preferencia a Direct3D apoyando de forma directa al uso de Windows, por lo que sus tarjetas siempre presentarán un mejor comportamiento es sistemas Windows. Por el contrario nVidia no sólo posee buenos drivers de Linux sino que optimiza su soporte para OpenGL con un rendimiento muy bueno, lo que hace que para Linux sean una opción mejor. Bajo éste criterio, se descartó la opción de adquirir tarjetas ATI y se optó por aquellas con chipset NVIDIA.

De las diferentes tarjetas posibles, se tuvo que elegir entre tarjetas con bus AGP (Accelerated Graphics Port) cuya especificación 1.0 dan velocidades de 133 Mhz (AGP 1X) y 266 Mhz (en AGP 2X), y aquellas con bus PCI-Express (anteriormente conocido por las siglas 3GIO, 3rd Generation I/O), un nuevo desarrollo del bus PCI que usa los conceptos de programación y los estándares de comunicación existentes, pero se basa en un sistema de comunicación serie mucho más rápido (apoyado principalmente por Intel).

En el momento de tomar la decisión, y debido a que la mayoría de los PC del laboratorio del IOC tenían bus AGP, se descartaron aquellas tarjetas con buses PCI-express. Sin embargo, el tiempo ha demostrado que la decisión, forzosamente tomada debido al HW del que se disponía, no fue la más acertada, ya que actualmente no se desarrollan mejoras sobre el puerto AGP, y éste, está siendo reemplazado por el bus PCI-Express (en el que entre otras mejoras, se pueden conectar más de una placa, obteniendo trabajo en paralelo para el procesamiento de video).

Finalmente, de entre las tarjetas restantes con buses AGP, se premió a aquellas que, teniendo soporte estéreo de tipo Quad buffer, tubiesen mejor relación calidad-precio. Así, finalmente, la opción elegida fue una tarjeta gráfica NVIDIA Quadro FX 1100.

7.3. Desarrollo del Software

7.3.1. Servidor de vídeo

El servidor de vídeo usado en este módulo es un *demonio* denominado SPOOK que permite capturar, procesar, codificar y distribuir flujos de vídeo y audio. Puede usarse entre otras aplicaciones para webcams, vigilancia de vídeo, videoconferencias, etc.

La entrada de vídeo puede alimentarse con dispositivos Video4Linux, cámaras IIDC Firewire; el audio, desde dispositivos OSS o ALSA; las codificaciones permitidas son entre otras, MPEG-4 via XviD, y JPEG via JPEGlib. Cabe destacar que flujos RTP codificados en MPEG-4 están disponibles vía RTSP.

Estas opciones se configuran en un fichero (spook.conf) que se divide en opciones globales y bloques de declaraciones. El conjunto de parámetros de opciones globales son el número de puerto para las conexiones entrantes. El bloque de declaraciones, por su lado, define la entradas, los filtros, los encoders y las salidas. Estos bloques definirán una función específica que será desempeñada por el flujo de vídeo y tendrá una entrada y una salida para esta función. La entrada de cada bloque coincidirá con la salida del bloque previo.

Debido a que en este proyecto, se tiene una capturadora de vídeo, no se definirá una entrada (input) porque obtendrá dicho flujo a través del harware de la tarjeta. Igualmente, por usarse RTSP para la transmisión de vídeo, no se definirá una salida (output) como bloque, por lo que se enviará a través de la red hacia los clientes.

El fichero de configuración para SPOOK utilizado para el proyecto se muestra en el anexo G.

7.3.2. Configuración de la conexión

En este apartado se describe la interacción entre el cliente programado para este módulo y el servidor *SPOOK* basado en RTSP.

Cuando el cliente desea iniciar una conexión con el servidor envía un comando del tipo OPTIONS `rtsp://servername:port/video` para conocer qué clase de métodos RTSP soporta el servidor. En este caso los métodos soportados por *SPOOK* son OPTIONS, DESCRIBE, PLAY, PAUSE y TEARDOWN, descritos de forma genérica a continuación:

OPTIONS Este comando permite conocer los métodos soportados por el servidor RTSP.

DESCRIBE Describe la información necesaria para configurar la transmisión, contiene datos como el tipo de flujo: audio o vídeo, el tipo de protocolo de transporte a usar, el tipo de codec, etc. La información devuelta por el servidor usa el formato del protocolo SDP.

SETUP Con este comando el cliente especifica el tipo de protocolo y el número de los puertos para la conexión. El servidor responde generando una sesión con un identificador para referirse a la conexión actual y asignando el número de los puertos a utilizar.

PLAY Este método indica al servidor que inicie el envío de datos con la información especificada en SETUP o, que reanude la transmisión después de un comando PAUSE.

PAUSE Pausa temporalmente la entrega de datos por parte del servidor.

TEARDOWN Detiene la entrega de datos y termina la sesión o conexión. Para volver a recibir datos después de este comando se requiere enviar de nuevo un comando SETUP.

Para la ejecución de estos comandos en el cliente se utiliza la librería *Live555* que ya tiene implementado los protocolos RTSP y RTP/RTCP.

Los detalles de programación se presentan en el anexo H.

7.3.3. Recepción, decodificación y visualización de datos

Para la obtención correcta de los dos flujos de datos que se pretenden recibir, es necesario un modelo concurrente, esto es, necesitan ejecutarse los dos flujos de forma paralela para ver los frames contenidos en tiempo real, aunque esta especificación puede no se cumplirse en transmisiones sobre Internet si se produce un retraso en las comunicaciones.

La recepción de datos se realiza mediante el uso de un hilo de programación, que está continuamente recibiendo información desde el servidor, a su vez, cada frame obtenido se decodifica y se guarda en una cola FIFO utilizada como buffer, asegurando de esta forma que el primer frame escrito sea el primer frame en mostrarse. Posteriormente, cada cierto tiempo y dependiendo del numero de frames por segundo, se extrae de la cola un frame y se visualiza. La figura 7.8 muestra el proceso descrito anteriormente.

Cabe destacar que la primera vez que se ejecuta el hilo se guardan varios frames en el buffer antes de iniciar la visualización.

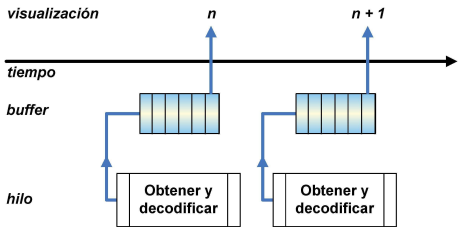

Figura 7.8 Flujo de recepción de datos de las imágenes

7.3.4. Librerías usadas en el módulo de visualización

Respecto a las librerías usadas en el módulo de visualización una vez que el vídeo ha sido transmitido, se ha usado, para el tratamiento del vídeo y su decodificación, libavcodec y libavformat y, para la ventana de visualización de las imágenes, OpenGL. Se presenta a continuación una descripción general de las mismas.

Por un lado, las librerías libavformat y libavcodec se han usado como base para la descompresión del formato MPEG-4 del flujo de vídeo. Éstas, pertenecen a su vez a una librería más extensa llamada ffmpeg, que permite el acceso a una gran variedad de formatos de vídeo. Aunque no existe una documentación propiamente dicha de la misma, existen diversas páginas web con programas de ejemplo que dan una idea general de cómo usarlas.

Se han usado estas dos librerías debido a su especialización en dos tareas diferentes. Esto es, muchos formatos de archivo de vídeo (AVI por ejemplo) a día de hoy no especifican qué codec usan para codificar la información de vídeo ni de audio; se limitan a definir cómo un flujo de audio y/o vídeo (o, potencialmente, diversos flujos audio/vídeo) se combinan en un mismo archivo. Este es el motivo por el que libavformat se encarga del problema de parsear los archivos de vídeo y separar los diferentes flujos contenidos en él, y sor su lado, libavcodec, de decodificar el flujo de imagen propiamente dicho.

Por otro lado, se ha usado OpenGL (Open Graphics Library), cuya traducción es biblioteca de gráficos abierta, por ser una especificación estándar que define una API multi-lenguaje multiplataforma para escribir aplicaciones que producen gráficos 3D. Desarrollada originalmente por Silicon Graphics Incorporated (SGI) destacan entre sus características, ser multiplataforma (habiendo incluso un openGL ES para móviles), y que permiten una gestión de la generación de gráficos 2D y 3D por hardware, ofreciendo al programador una API sencilla, estable y compacta. Además su escalabilidad ha permitido que no se estanque su desarrollo, permitiendo la creación de extensiones, una serie de añadidos sobre las funcionalidades básicas, en aras de aprovechar las crecientes evoluciones tecnológicas.

Igualmente, se han incluido en el proyecto las funcionalidades de GLUT (OpenGL Utility Toolkit), una sencilla API multiplataforma que provee funcionalidades para el manejo de ventanas e interacción por medio de teclado y ratón. Su utilidad principal es permitir el desarrollo de aplicaciones con OpenGL sin demasiado esfuerzo, dejando a un lado la complejidad de otros sistemas de ventanas, sin dependencia del sistema operativo, lo que permite una fácil migración de una plataforma a otra. Actualmente existen varias implementaciones de GLUT, como el proyecto Open Source FREEGLUT.

7.3.5. Particularidades del software de visualización

Aunque los pormenores de la programación de éste módulo están contenidos en la documentación de la interfaz anexada en H a esta memoria, se muestran a continuación las particularidades del mismo:

Posibilidad de intercambio entre estéreo y mono

Para el desarrollo del software, la interfaz diseñada, al ser previsto su uso en diferentes aplicaciones de teleoperación y control de la celda robotizada del laboratorio, se ha tenido en cuenta que el módulo debe poder beneficiarse de una dualidad entre vídeo monoscópico y estereoscopía, por ello, la interfaz debe contemplar la posibilidad de cambiar como mínimo de una visión estéreo a una monoscópica en el momento de arrancar la aplicación.

Para lograrlo, se han definido varios modos distintos de visualización: el modo MONOVIEW, en el que sólo se visualiza un vídeo monoscópico; un modo STEREO, en el que actúa el quad buffer y se permite la visión estereoscópica de pares de imágenes, y finalmente, un tercer modo ALIGN que permite ver separadamente dos imágenes procedentes de dos vídeos diferentes que servirá para poder realizar pruebas con las imágenes.

Posibilidad de visualización

En el tiempo de desarrollo de este proyecto, se ha implementado un programa que permite visionar una imagen, un vídeo monoscópico, un vídeo transmitido por Internet procedente de las cámaras vía RTSP, la visualización tridimensional de pares estereoscópicos, la visualización de dos vídeos pre-grabados. Actualmente, para completar las funcionalidades de este programa, faltaría integrar la posibilidad de visualización de dos vídeos que formen par estereoscópico de manera sincronizada, que está actualmente en desarrollo en el IOC.

La elección de uno u otro modos de visión se lleva a cabo mediante banderas al arrancar el programa. El aspecto de la interfaz de visualización tridimensional se muestra en la figura 7.9.

Los pormenores de la programación del visor, se detallan en el anexo H.

Figura 7.9 Foto Real de la interfaz

8 Análisis experimental y Resultados

Después de haber implementado el sistema descrito, se ha procedido a su puesta en marcha y a la realización de diversos experimentos para su validación y mejora.

8.1. Experimentación con el módulo sensor

8.1.1. Configuración de las cámaras

Después de implantar el módulo sensor, se constata experimentalmente que por limitaciones físicas de la configuración de las cámaras, cuando sus ejes están en paralelo y se hallan situadas a la menor distancia posible la una de la otra, se restringen entre ellas el campo de visión, por lo que para trabajos futuros, se propone terminar la implemantación del cliente del sensor imagen incluyendo entre sus funcionalidades la sincronización de los movimientos de las mismas para que enfoquen un mismo punto elegido por el usuario.

8.1.2. Calidad de los pares estereoscópicos

Se ha constatado también que, al situar las dos cámaras VC-C5i, supuestamente calibradas por el fabricante, en una superficie vertical con sus ejes en paralelo y en su posición inicial, es decir, con los ángulos de giro de sus ejes a cero y sin zoom, se aprecia una disparidad vertical no despreciable entre las imágenes obtenidas. Esto es debido al huelgo mecánico de sus ejes de giro, lo que supone que no se pueda asegurar sin un calibrado por sofware que los pares esteroscópicos que se obtengan de éstas, se perciban por el usuario como imágenes con percepción de profundidad cuando se transmitan vía RTSP hacia el módulo de visualización.

Para paliar este problema, se propone como trabajo futuro, el uso de algunos de los múltiples algoritmos de calibración para cámaras con tres grados de libertad, como el de Zhang [37], método de calibración en el que sólo conociendo la distancia entre las marcas de un patrón de calibración y a partir de diferentes vistas de dicho patrón, es capaz de hallar los parámetros de calibración de las cámaras; el de Heikkila [38], que usa técnicas de minimización de mínimos cuadrados y necesita el conocimiento de las parcas del patrón en el sistema de referencia global y

su correspondencia con las coordenadas de las cámaras; u otros frecuentemente utilizados como el método de calibración de Tsai [31] y el de Faugeras [11].

Cabe destacar, que por sus características según la bibliografía consultada se recomienda el uso del método de Zhang o, como alternativa el algoritmo desarrollado por Juan Andrade Cetto, ya que en ambos casos su implementación es relativamente sencilla y permiten obtener buenos resultados.

8.2. Experimentación de la sensación estereoscópica

Para el análisis de la percepción estereoscópica de los usuarios del sistema se han capturado doce imágenes estáticas de una misma escena de la celda robotizada, con diferentes disparidades, llegando a los siguientes resultados.

Con paralax horizontales ideales (cercanos a 6,5 cm), todos los usuarios a los que se les ha realizado un test, percibían las imágenes como tridimensionales. Para paralax superiores, no todos los usuarios perciben la sensación estéreo de modo distinto, lo que demuestra, como se ha mencionado anteriormente, que la percepción de profundidad es subjetiva, depende del usuario, y mejora con una mayor exposición a imágenes tridimensionales generadas en el sistema.

Se presenta en la figura 8.1, la imágenes usadas para el test de profundidad.

Figura 8.1 Vista del portátil de O.C. con diferentes valores de paralax

Gracias a esta experimentación, se llega a la conclusión que la percepción estereoscópica es subjetiva y mejora con la exposición prolongada a dicha sensación.

Existe una distancia ideal a la que todos los usuarios declararon haber obtenido la sensación de profundidad a partir de dos imágenes recogidas por las cámaras. Ésta se halla entre 6,5 y 7,5

de separación entre las dos muestras tomadas, aunque una gran la mayoría, seguía percibiendo sensación estereoscópica a distancias mayores. En ellas, aunque con sensación esteroscópica, la fatiga visual se veía incrementada.

8.3. Experimentación del comportamiento del zoom

Una vez implementada la corrección de la longitud focal en el módulo sensor, detallada en el apartado 6.2.4, y con los resultados experimentales presentados en el anexo J, se comprueba el efecto consegido con la extracción de la imágenes generadas por las dos cámaras en configuración de ejes ópticos paralelos.

Las visualización de las imágenes obtenidas por éstas indican que, para números de zoom pequeños, la solución implementada de modificación del zoom para la obtención del mismo tamaño de los objetos observados, cumple con las relaciónes de tamaños pretendidas. Sin embargo, para relaciones de zoom mayores, el comportamiento del sistema es inadecuado debido a la influencia de distiontos factores que se presentan a continuación.

El factor más influyente para el error en las medidas de corrección del zoom, se produce por el hecho que la curva de regresión que más se asemeja a los resultado experimentales sea logarítmica, ya que para longitudes focales crecientes, el error se hace cada vez mayor. Esto implicará que pequeñas variaciones en dicha longitud, producirá un incremento excesivo del factor de zoom.

En segundo lugar, el error se debe a las suposición, no realista, de que la lente motorizada de la cámara se comporta como en la ecuación de Gauss de las lentes delgadas.

Finalmente, y según el modelo de cámaras utilizado, el hecho de que las cámaras tengan los ejes paralelos, presupone que el observador debe mirar hacia el infinito, por lo que querer incrementar el zoom más de unas pocas unidades, indicaría que el observador quiere focalizar su atención en un punto en concreto de un objeto, y no en la supervisión de la escena en sí. Esta suposición, llevaría a ls cámaras a tener que producir una convergencia de ejes en dicho punto, para la que el modelo de configuración de las cámaras presentado en este proyecto no sería válido, y la implementación de dicho comportamiento quedaría por desarrollar en trabajos futuros.

9 Conclusiones y trabajos futuros

Una vez integrados los módulos presentados en este proyecto y analizados los resultados que se constata la consecución de los objetivos que se presentan a continuación.

En el módulo sensor se ha conseguido desarrollar un servidor específico para las cámaras Canon VC-C5 que permite el control remoto de las mismas. Se ha diseñado una arquitectura con modelo cliente-servidor con una librería que capaz de controlar tanto el movimiento de las dos cámaras a la vez, como su movimiento individualizado. Se ha diseñado e implementado una interfaz de usuario intuitiva capaz del control por software del movimiento de las cámaras cuando éstas tienen la configuración de ejes paralelos e implementado un *user interface manager* que facilita las tareas de supervisión de un programador.

En el módulo de visualización se ha alcanzado el objetivo de representar imágenes planas en dos formatos distintos como pares estereoscópicos con su correspondiente posibilidad de visualización tridimensional, se ha implementado una clase decodificadora que permite la reproducción de vídeos pregrabados en formato .avi y su reproducción dentro de la interfaz. Se ha logrado también integrar la retransmisión en tiempo real (sólo con el retraso propio de la red) un flujo de vídeo, la decodificación del mismo, y su representación correspondiente dentro de viewer3D.

Queda por añadir una lista de posibles trabajos futuros para continuar con el desarrollo de los dos módulos presentados, haciendo énfasis en los siguientes puntos de mejora. Por un lado, y con respecto al módulo sensor, se propone mejorar el servidor de las cámaras para obtener un mejor control de movimientos de las mismas; la implementación del segundo modelo de configuración presentado, cámaras con ejes convergentes, para ampliar el rango efectivo de visión de la celda robótica y la mejora del control de la óptica de la cámara [35]. Con respecto al módulo de visualización, se plantea la mejora de la interfaz de visión con la inclusión de un segundo flujo de imágenes vía RTSP que permitirá obtener los pares estereoscópicos deseados; implementar un algoritmo de sincronización temporal de flujos entre los buffers de imágenes.

Finalmente, se plantea como último objetivo, la incorporación de técnicas de realidad aumentada y la integración propiamente dicha de ambos módulos dentro del proyecto TASTRI2.

10 Agradecimientos

No quisiera acabar este proyecto sin antes dar las gracias a todas aquellas personas que me han apoyado en este proyecto, ya que sin su ayuda este proyecto no hubiera sido posible.

En primer lugar, quiero agradecer a todas aquellas personas que desarrollan desinteresadamente programas bajo licencias públicas y que comparten su conocimiento con los demás.

A los profesores, investigadores y becarios del Departamento del Instituto de Organización y Control de Sistemas Industriales (IOC) de la UPC por las múltiples ayudas prestadas, por apoyarme y aconsejarme cuando las cosas se han complicado y por acogerme entre ellos como uno más.

A mis antiguos compañeros de trabajo del Centro de Realidad Virtual, por su asesoramiento en dispositivos de Realidad Virtual y en la bibliografía, a Eva M. del Instituto de Robótica e Informática Industrial, por su tiempo y sus consejos.

A mis compañeros de trabajo de Servicios de Ingeniería y Packaging y de LSI, por su soporte moral.

A mis amigos, a aquellos que desde el principio han estado ahí para lo bueno y para lo malo, y a los que espero corresponder en algún momento.

Y finalmente a mi familia, en especial a mi padre y a mi hermana, por su paciencia durante todos estos meses y por el soporte incondicional que me han prestado, aunque en muchos momentos no haya sabido agradecerlo.

A todos vosotros, GRACIAS!.

Bibliografía

[1] Basel convention de las naciones unidas. *http://www.basel.int/.*

[2] Real decreto 208/2005, de 25 de febrero, sobre aparatos eléctricos y electró nicos y la gestió n de sus residuos. *http://www.boe.es.*

[3] Wastes from electrical & electronic equipment. *Directiva de la Unió n Europea, http://www.dti.gov.uk/sustainability/weee/index.htm.*

[4] Perry M. Agarwal, D. Camera remote control command language. *LBNL PUB-3149*, 1998.

[5] Tan W y Wee S. Apostoulopolous, J. Video streaming: Concepts, algorithms, and systems. *Reporte Tecnico HPL-2002-260 Mobile and Media Systems Laboratory HP Laboratories Palo Alto*, 2002.

[6] L. Byster, S. Westervelt, R. Gutierrez, and A.and Dutta M. Davis, S.and Hussain. Exporting harm: The high-tech trashing of asia. *Basel Action Network & Silicon Valley Toxics Coallition*, 2002.

[7] Canon. Vc-c5i comunication camera. programmer's manual ver 1.1. *http://www.usa.canon.com/consumer/controller*, 5 Enero 2005.

[8] N. Chang and A. Zakhor. View generation for three-dimensional scenes from video sequences. *IEEE Trans. Image Process.*, vol. 6, pp. 584?598.

[9] Handy & Harman Electronic Materials Corp. Tabla presentada en microelectronics and computer technology corporation 1996. *http://www.handyharman.com/.*

[10] Neil A. Dodgson. Autostereoscopic 3d displays. *University of Cambridge Computer Laboratory.*

[11] O. Faugeras. Three-dimensional computer vision: A geometric viewpoint. *Cambridge, MA: MIT Press*, 1993.

[12] Aracil R. Navas M Ferrer, M. Procesamiento de vídeo en tiempo real para teleoperació n: superposició n de imágenes y estereoscopía. *UPM.*

[13] R. Franich. *Tisparity estimation in stereoscopic digital images.* PhD thesis, 1996.

[14] J.P. Frisby. Stereo and texture cue integration in the perception of planar and curved large real surfaces. in attention and performance 16: Information integration in perception and communication. *In T. Inui & J. L. McClelland (Eds.), Attention and performance 16: Information integration in perception and communication (pp. 71-92). Cambridge, MA: MIT Press.*, 1996.

[15] M. J. García. *Medi ambient i tecnologia*. Edicions UPC, 1998.

[16] Möller M. C. & Wensveen J. M. Harwerth, R. S. Effects of cue context on the perception of depth from combined disparity and perspectuve cues. *Optometry and Vision Science, 75, 433-444.*, 1998.

[17] Seuntiens P.J.H. IJsselsteijn, W.A. and L.M.J. Meesters. State-of-the-art in human factors and quality issues of stereoscopic broadcast television. *Deliverable ATTEST/WP5/01 IST-2001-34396*, August 2002.

[18] Jonathan G. Koomey Bruce Nordman Richard E. Brown Mary Ann Piette Michael Ting Kawamoto, Kaoru and Alan K. Meier. *Electricity Used by Office Equipment and Network Equipment in the U.S.: Detailed Report and Appendices.* 2001.

[19] W. S Kim. Virtual reality calibration and predective displays for telerobotics. *Presence, MIT Press Cambridge*, 5(2):173–190, 1983.

[20] D. L. Mills. Network time protocol (version3) specification, implementation and analysis. *RFC1305*, 1992.

[21] Revista Opcions nº6. *Els Ordinadors.* Number 6. 2003.

[22] Mallem M. Kheddar A. Otmane, S. and F. Chavand. Active virtual guides as an apparatus for augmented reality based telemanipulation system on the internet. *Proc, 33rd Annual Simulation Symposium, 2000. (SS 2000) 16-20 April 2000 p 185-191.*, 2001.

[23] M. Perry and D. Agarwal. Remote control for videoconferencing. *Information and Computing Sciences Division Ernest Orlando Lawrence Berkeley National Laboratory*, 1998.

[24] Jim Puckett. The real-life recycling horror show. *http://www.ban.org/Library/Jim %20Puckett %27s %20Guest %20Column.pdf.*

[25] Kurt W. Roth, F. Goldstein, and J. Kleinman. *Energy Consumption by Office and Telecommunications Equipment in Commercial Buildings*, volume Energy Consumption Baseline. 2002.

[26] W. Fred Goldstein Roth, Kurt and Jonathan Kleinman. *Energy Consumption by Office and Telecommunications Equipment in Commercial Buildings*, volume I. Energy Consumption Baseline, 2002.

[27] M Sanchez Moreno. *Reconstrucció n Tridimensional de Escenas con Iluminació n Láser: Aplicaciones a la Fotogrametría Industrial*. PhD thesis, 2000.

[28] R. Schulzrinne, H. y Rao. Real-time streaming protocol. *RFC2326 Internet Engineering Task Force*, 1998.

[29] V Sonka, M. Hlavac and R. Boyle. *Image processing, Analysis and MAchine Vision.* 1998 Second Edition, ISBN 0-534-95393-X.

[30] Michael W. Toffel and A. Horvath. Environmental implications of wireless technologies: News delivery and business meetings.

[31] R. Tsai. An efficient and accurate camera calibration technique for 3d machine vision. *Robotics and Automation, IEEE Journal of [legacy, pre - 1988], Volume: 3, Issue: 4 páginas: 323- 344, ISSN: 0882-4967*, Aug 1987.

[32] Eric D. Williams. Revisiting energy used to manufacture a desktop computer: hybrid analysis combining process and economic input-output methods.

[33] Eric D. Williams. *Environmental impacts of microchip manufacture*, volume 461. August 2004.

[34] Eric D Williams, Robert U. Ayre, and M. Heller. The 1.7 kilogram microchip: Energy and material use in the production of semiconductor devices. *http://pubs.acs.org/cgi-bin/article.cgi/esthag/2002/36/i24/pdf/es025643o.pdf.*

[35] G. Willson. Modeling and calibration of automated zoom lenses. *http://vasc.ri.cmu.edu/IUS/usrp2/rgw/www/spie94.pdf*, 1994.

[36] Docherty T. Woods, A. and R. Koch. Image distortions in stereoscopic video systems. *in Proc. vol. 1915, SPIE Stereoscopic Displays and Applications*, Feb. 1993.

[37] Z. Xu, G.; Zhang. Epipolar geometry in stereo, motion and object recognition: A unified approach. *Kluwer Academic Publishers, Dordrecht, Boston, London,*, 1996.

[38] Janne Heikkila y Olli Silven. A four-step camera calibration procedure with implicit image correction. *En Proceedings of the IEEE Conference on Comput ter Vision and Pattern Recognition*, 1996.

Bibliografía

[1] Basel convention de las naciones unidas. *http://www.basel.int/.*

[2] Real decreto 208/2005, de 25 de febrero, sobre aparatos eléctricos y electró nicos y la gestió n de sus residuos. *http://www.boe.es.*

[3] Wastes from electrical & electronic equipment. *Directiva de la Unió n Europea, http://www.dti.gov.uk/sustainability/weee/index.htm.*

[4] Perry M. Agarwal, D. Camera remote control command language. *LBNL PUB-3149*, 1998.

[5] Tan W y Wee S. Apostoulopolous, J. Video streaming: Concepts, algorithms, and systems. *Reporte Tecnico HPL-2002-260 Mobile and Media Systems Laboratory HP Laboratories Palo Alto*, 2002.

[6] L. Byster, S. Westervelt, R. Gutierrez, and A.and Dutta M. Davis, S.and Hussain. Exporting harm: The high-tech trashing of asia. *Basel Action Network & Silicon Valley Toxics Coallition*, 2002.

[7] Canon. Vc-c5i comunication camera. programmer's manual ver 1.1. *http://www.usa.canon.com/consumer/controller*, 5 Enero 2005.

[8] N. Chang and A. Zakhor. View generation for three-dimensional scenes from video sequences. *IEEE Trans. Image Process.*, vol. 6, pp. 584?598.

[9] Handy & Harman Electronic Materials Corp. Tabla presentada en microelectronics and computer technology corporation 1996. *http://www.handyharman.com/.*

[10] Neil A. Dodgson. Autostereoscopic 3d displays. *University of Cambridge Computer Laboratory.*

[11] O. Faugeras. Three-dimensional computer vision: A geometric viewpoint. *Cambridge, MA: MIT Press*, 1993.

[12] Aracil R. Navas M Ferrer, M. Procesamiento de vídeo en tiempo real para teleoperació n: superposició n de imágenes y estereoscopía. *UPM.*

[13] R. Franich. *Tisparity estimation in stereoscopic digital images.* PhD thesis, 1996.

[14] J.P. Frisby. Stereo and texture cue integration in the perception of planar and curved large real surfaces. in attention and performance 16: Information integration in perception and communication. *In T. Inui & J. L. McClelland (Eds.), Attention and performance 16: Information integration in perception and communication (pp. 71-92). Cambridge, MA: MIT Press.*, 1996.

[15] M. J. García. *Medi ambient i tecnologia*. Edicions UPC, 1998.

[16] Möller M. C. & Wensveen J. M. Harwerth, R. S. Effects of cue context on the perception of depth from combined disparity and perspectuve cues. *Optometry and Vision Science, 75, 433-444.*, 1998.

[17] Seuntiens P.J.H. IJsselsteijn, W.A. and L.M.J. Meesters. State-of-the-art in human factors and quality issues of stereoscopic broadcast television. *Deliverable ATTEST/WP5/01 IST-2001-34396*, August 2002.

[18] Jonathan G. Koomey Bruce Nordman Richard E. Brown Mary Ann Piette Michael Ting Kawamoto, Kaoru and Alan K. Meier. *Electricity Used by Office Equipment and Network Equipment in the U.S.: Detailed Report and Appendices.* 2001.

[19] W. S Kim. Virtual reality calibration and predective displays for telerobotics. *Presence, MIT Press Cambridge*, 5(2):173–190, 1983.

[20] D. L. Mills. Network time protocol (version3) specification, implementation and analysis. *RFC1305*, 1992.

[21] Revista Opcions nº6. *Els Ordinadors.* Number 6. 2003.

[22] Mallem M. Kheddar A. Otmane, S. and F. Chavand. Active virtual guides as an apparatus for augmented reality based telemanipulation system on the internet. *Proc, 33rd Annual Simulation Symposium, 2000. (SS 2000) 16-20 April 2000 p 185-191.*, 2001.

[23] M. Perry and D. Agarwal. Remote control for videoconferencing. *Information and Computing Sciences Division Ernest Orlando Lawrence Berkeley National Laboratory*, 1998.

[24] Jim Puckett. The real-life recycling horror show. *http://www.ban.org/Library/Jim %20Puckett %27s %20Guest %20Column.pdf.*

[25] Kurt W. Roth, F. Goldstein, and J. Kleinman. *Energy Consumption by Office and Telecommunications Equipment in Commercial Buildings*, volume Energy Consumption Baseline. 2002.

[26] W. Fred Goldstein Roth, Kurt and Jonathan Kleinman. *Energy Consumption by Office and Telecommunications Equipment in Commercial Buildings*, volume I. Energy Consumption Baseline, 2002.

[27] M Sanchez Moreno. *Reconstrucció n Tridimensional de Escenas con Iluminació n Láser: Aplicaciones a la Fotogrametría Industrial.* PhD thesis, 2000.

[28] R. Schulzrinne, H. y Rao. Real-time streaming protocol. *RFC2326 Internet Engineering Task Force*, 1998.

[29] V Sonka, M. Hlavac and R. Boyle. *Image processing, Analysis and MAchine Vision.* 1998 Second Edition, ISBN 0-534-95393-X.

[30] Michael W. Toffel and A. Horvath. Environmental implications of wireless technologies: News delivery and business meetings.

[31] R. Tsai. An efficient and accurate camera calibration technique for 3d machine vision. *Robotics and Automation, IEEE Journal of [legacy, pre - 1988], Volume: 3, Issue: 4 páginas: 323- 344, ISSN: 0882-4967*, Aug 1987.

[32] Eric D. Williams. Revisiting energy used to manufacture a desktop computer: hybrid analysis combining process and economic input-output methods.

[33] Eric D. Williams. *Environmental impacts of microchip manufacture*, volume 461. August 2004.

[34] Eric D Williams, Robert U. Ayre, and M. Heller. The 1.7 kilogram microchip: Energy and material use in the production of semiconductor devices. *http://pubs.acs.org/cgi-bin/article.cgi/esthag/2002/36/i24/pdf/es025643o.pdf.*

[35] G. Willson. Modeling and calibration of automated zoom lenses. *http://vasc.ri.cmu.edu/IUS/usrp2/rgw/www/spie94.pdf*, 1994.

[36] Docherty T. Woods, A. and R. Koch. Image distortions in stereoscopic video systems. *in Proc. vol. 1915, SPIE Stereoscopic Displays and Applications*, Feb. 1993.

[37] Z. Xu, G.; Zhang. Epipolar geometry in stereo, motion and object recognition: A unified approach. *Kluwer Academic Publishers, Dordrecht, Boston, London,*, 1996.

[38] Janne Heikkila y Olli Silven. A four-step camera calibration procedure with implicit image correction. *En Proceedings of the IEEE Conference on Comput ter Vision and Pattern Recognition*, 1996.